THE UNIVERSITY *of* EDINBURGH

Structural and Functional Interrogation of Anterior Gradient-2

Terry A. Gray

Thesis submitted for the degree of Doctor of Philosophy

University of Edinburgh

2013

Contents

Figures and Tables

Acknowledgements

First of all, Thanks to Prof Ted,

Who never seems to wake up on the wrong side of bed,

Plenty of experiment ideas to do,

All to please the peer review.

Can't forget Dr Euan Murray,

For methods, tips and the occasional curry.

His World Cup knowledge will never be beat,

Just don't sit in his tea-time seat.

Collaborating scientists over in Czech,

Borek ever the host, with wine for our neck!

Without any of you, this could never be done,

I'm grateful to everyone, who've made it so fun.

Magda, Vivien, Jenny and Jude.

Friendly faces, good chat to lighten the mood.

Add to that all the people of labs Hupp and Ball,

To many to mention, so 'A huge Thank You' to all.

The family next, John, Mum and Dad.

Supportive encouraging though good times and bad.

Teasing and joking of the eternal student,

Helpful with finances when being less prudent.

Grandad, Gran and Auntie Dot,

Gifting me alcohol and food whether they like it or not!

The lads, Gavin, Calum, Jordan and Ben,

Nights out and sport, to help keen me sane.

Lauren, my pal, with her wee friendly face.

Kicking my ass, so I know my place.

Finally, Susan my soul mate and friend,

Always ready with a patient listening ear to lend.

I’ve surprised myself and ended up here

With a thesis in both hands, and soon hopefully beer.

These several years, I will never forget.

Who knows where, science will take me next...

Declaration

I hereby declare that I am the author of this thesis. The work herein is entirely my own unless otherwise clearly indicated and acknowledged. I can confirm that this thesis has been submitted for the degree of Doctor of philosophy and no part of this work has been submitted for any other degree or professional qualification.

Abbreviations

(H)/KDEL	(Histidine)/Lysine-Aspartate-Glutamate-Leucine canonical ER retention motif
(v/v)	volume to volume
(w/v)	weight to volume
17-AAG	17-allylamino-17-demethoxygeldanamycin
2-DIGE	2-D Fluorescence Difference Gel Electrophoresis
^{2S}MTA	Two-site sandwich microtitre assay
A	Adenine
A	Alanine (Ala)
AAA+	ATPases associated with diverse cellular activities
AGR2	Anterior Gradient 2
AGR2-KDEL	Mutation of ER-retention sequence of AGR2 to optimal
AGR2-Δ45	Removal of N-terminal 45 amino acids of AGR2
AGR2-ΔC	Removal of C-terminal ER retention motif of AGR2
AGR3	Anterior Gradient 3
AQUA	Absolute quantitation
AREG	Amphiregulin
ATG	Adenine-Thymine-Guanine start codon
ATP	Adenosine Triphosphate
b	Nucleotide bases
kb	Nucleotide kilobases
Bak	Bcl-2 homologous antagonist/killer
BAX	Bcl-2–associated X protein
Bcl-2	B-cell lymphoma-2
BCL-x	Bcl-2-like protein 1
BCR-Abl	Breakpoint cluster region-Abelson fusion protein
BGH	bovine growth hormone
BID	BH3 interacting-domain death agonist
BiP/GRP78	Binding immunoglobulin protein/78 kDa glucose-regulated protein
BSA	Bovine serum albumin
C	Cytosine
C	Cysteine (Cys)
C4.4	LYPD3 - LY6/PLAUR domain containing 3
CALU	calumenin
cdk	Cyclin-dependent kinase
CDK2	Cyclin dependent kinase 2
cDNA	copy DNA

CHOP	CCAAT/-enhancer-binding protein homologous protein
CID	collision inducted dissociation
CO_2	Carbon dioxide
CoIP	Co-immunoprecipitation
Ct	cycle threshold
C-terminal	Carboxyl (-COOH) terminus of polypeptide/protein
CXXC	Cysteine-X-X-Cysteine Thioredoxin fold
D	Aspartic Acid (Asp)
Da	Daltons
DAVID	Database for Annotation, Visualisation and Integrated Discovery
DMEM	Dulbecco's modified eagle medium
DMSO	Dimethylsulphoxide
DNA	Deoxyribonucleic acid
dNTP	deoxyribonucleotides
DSS	Disuccinimidyl suberate
DTT	Dithiothreitol
DyL800	DyLight 800 fluorophore
E	Glutamic Acid (Glu)
E. coli	*Escherichia coli*
E2F1	Transcription factor E2F1
EBP-1	Receptor tyrosine-protein kinase ErbB-3 binding protein 1
ECL	Enhanced Chemiluminescence
EDTA	Ethylenediaminetetraacetic acid
EGF	Epidermal growth factor
EGFR	epidermal growth factor receptor
EGTA	ethylene glycol tetraacetic acid
ELISA	Enzyme-linked immunosorbent assay
ER	Endoplasmic reticulum
ERK1/2	extracellular-signal-regulated kinases 1/2
ERP	Endoplasmic reticulum protein
ES	Electrospray
ESCRT-1	Endosomal Sorting Complexes Required for Transport
F	Phenylalanine (Phe)
FACS	Fluorescence activated cell sorting
FBS	Foetal bovine serum
FITC	Fluorescein isothiocyanate
FOXA1/FOXA2	forkhead box transcription factors
FRT	Flippase recombination target
G	Guanine
g	grams

G	Glycine (Gly)
G1	Gap phase 1 of cell cycle
G1/S	Gap 1 - Synthesis transition of cell cycle
G2/M	Gap phase 2-Mitosis transition DNA damage checkpoint of the cell cycle
GADD	Growth Arrest and DNA Damage Proteins
GAGE	G-antigen
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
GFP	green fluorescent protein
GOI	gene of interest
GPI	glycophosphatidylinositol
GST	Glutathione *S*-transferase
h	hydrophobic residue
H	Histidine (His)
HCl	Hydrochloric acid
HER2/neu	Human epidermal growth factor 2 receptor
HIF-1	hypoxia inducible factor 1α
HIP2	Ubiquitin-conjugating enzyme E2 K
HPLC	High-performance liquid chromatography
HRP	horseradish peroxidase
Hsp70	Heat-shock protein 70 kDa
Hsp90	Heat-shock protein 90 kDa
I	Isoleucine (Ile)
ICAT	isotope-coded affinity tag
IL-13	Interleukin-13
IMAC	Immobilized metal ion affinity chromatography
IPA	Ingenuity pathway analysis
IPI	international protein index
IPTG	Isopropyl β-D-1-thiogalactopyranoside
IR	Infra-red
IRE1-α	Serine/threonine-protein kinase/endoribonuclease
IRF-1	Interfereon regulatory factor-1
ITRAQ	isotope tags for relative and absolute quantification
JNK	c-Jun N-terminal kinases
K	Lysine (Lys)
kb	kilo-bases
Kd	Dissociation constant
kDa	Kilo-Daltons
KEGG	Kyoto Encyclopaedia of Genes and Genomes
Ki-67	Antigen KI-67
KTEL	Lysine-Threonine-Glutamate-Leucine putative ER retention motif

L	Litre
L	Leucine (Leu)
LB	Luria Bertani Broth
LC	Liquid chromatography
M	Molar
M	Methionine (Met)
m/z	mass/charge
mA	milliamperes
MAB	Monoclonal antibody
mAGR2/GOB4	Murine Anterior Gradient 2
MALDI	Matrix-associated laser desorption ionisation
MCS	Multiple cloning site
MDM2	Mouse double minute 2
mg	milligrams
mL	millilitre
mM	millimolar
mm	millimetre
MOMP	mitochondrial outer membrane permeabilisation
mRNA	Messenger RNA
MS	Mass spectrometry
MS/MS	tandem mass spectrometry
MTT	tetrazolium dye
MUC1	Mucin 1
MUC2	Mucin 2
MUC5AC	Mucin 5AC
MUC5B	Mucin 5B
N	Asparagine (Asn)
nAG	Newt Anterior Gradient gene
NEB	New England Biolabs
NF-κB	nuclear factor kappa-light-chain-enhancer of activated B cells
ng	nanograms
NHS	*N*-Hydroxysuccinimide
nL	nanolitre
nm	nanometre
NMR	Nuclear magnetic resonance
NOXA	Phorbol-12-myristate-13-acetate-induced protein 1
N-terminal	Amino (-NH2) terminus of polypeptide/protein
P	Proline (Pro)
p21-(WAF-1)	Cyclin-dependent kinase inhibitor 1 (wild-type p53 activated fragment 1)
p38	p38 mitogen-activated protein kinases

p53	Tumour protein 53
PaCIFIC	Precursor Acquisition Independent From Ion Count
PBS	phosphate buffered saline
PBST	PBS supplemented with 0.05% Tween-20
PCR	Polymerase chain reaction
PDI	Protein disulphide isomerase
PDPK1-AKT	phosphoinositidine-dependent protein kinase 1-AKT
PEP	posterior error probability
PERK	protein kinase RNA-like endoplasmic reticulum kinase
pI	Isoelectric point
PI3K	Phosphatidylinositide 3-kinase
Poly(I):poly(C)	polyinosinic polycytidylic acid
polyA	poly-adenosine tail
PRIMA-1	p53 re-activation and induction of massive apoptosis-1
Prod-1	Axolotl homolog for human CD59
PSI	Pounds per square inch
PSMA3	Proteasome subunit alpha type-3
PSMB2	Proteasome subunit beta type-2
PSMB4	Proteasome subunit beta type-4
PSMC3	26S protease regulatory subunit 7
PUMA	p53-upregulated modulator of apoptosis
Q	Glutamine (Gln)
qPCR	quantitative Real-time PCR
R	Arginine (Arg)
rcf	Relative centrifugal force
RCN1	reticulocalbin 1
RIN	RNA integrity number
RIPA buffer	Radioimmunoprecipitation assay buffer
RITA	reactivation of p53 and induction of tumour cell apoptosis
RNA	ribonucleic acid
RNAi	RNA interference
rpm	revolutions per minute
RPMI	Roswell Park Memorial Institute
RT	Reverse transcription
RTCA-(DP)	Real-time cell analyzer-(dual plate)
S	Serine (Ser)
SDS-PAGE	Sodium dodecyl sulphate-polyacrylamide gel electrophoresis
SEC	size exclusion chromatography
SEC-MALLS	size exclusion chromatography-multi angle laser light scattering

SERCA	sarco/endoplasmic reticulum Ca^{2+}-ATPase
SERS	Surface enhanced Raman scattering
SIDR	Strathclyde Innovations in Drug Research
SILAC	Stable isotope labelling by amino acids in cell culture
siRNA	short interfering RNA
SMAD4	SMAD family member 4
SNP	Single nucleotide polymorphism
SOX9	Transcription factor SOX-9
SRP	Signal recognition particle
SV40	Simian Vacuolating Virus 40
T	Thymine
T	Threonine (Thr)
TAE	Tris base, Acetic acid, EDTA
TALEN	transcription activator-like effector nuclease
TBS	Tris buffered saline
TCDD	2,3,7,8-Tetrachlorodibenzo-p-dioxin
TDA	Thermal denaturation assay
TEV	Tobacco etch virus
TGF-β	tumour growth factor-β
TLR3	toll-like receptor 3
TMT	Tandem mass tags
TSG101	Tumour susceptibility gene 101
Tween-20	Polyoxyethylene sorbitan monolaurate
TX	Thioredoxin fold
U	units
Ub	Ubiquitin
Ubc	ubiquitin conjugating domain
UPR	Unfolded protein response
UV	Ultra-violet
V	Volts
V	Valine (Val)
VEGF-A	vascular endothelial growth factor-A
W	Tryptophan (Trp)
WT	Wild-type
x	any amino acid residue
XAG	*Xenopus laevis* Anterior Gradient gene
Y	Tyrosine (Tyr)
Y2H	Yeast-2-Hybrid
Yap-1	Yes-associated protein-1
z-AG2	Zebrafish (*Dario renio*) Anterior Gradient gene
ZNF217	Zinc finger protein 217

α-DAG	alpha-dystroglycan
µg	micrograms
µL	microlitre
µM	micromolar

Abstract

Anterior Gradient-2 protein (AGR2) has recently been linked to the onset of several pathologies including asthma and inflammatory bowel disease. Most interestingly, it has been discovered to influence the transformation of cells and metastatic growth essential to cancer development, and has subsequently been linked to the development of resistance to anti-cancer therapeutics. AGR2 protein is overexpressed in a diverse range of human cancer types, and has been detected secreted into the extracellular milieu. Thus, AGR2 protein represents a compelling pro-oncogenic signalling intermediate in tumour emergence and endurance. This thesis presents an interdisciplinary approach including structural biology, cell biology and synthetic biology, and clinical studies to shed more light on the role of AGR2 in cancer development. Synthetic cell based reagents were developed to define the dominant pathways that are reprogrammed in a cell as a result of AGR2 synthesis. A cell panel was engineered incorporating the AGR2 (and mutants thereof) allele into the AGR2-null A375 cell line. These tools were then coupled to quantitative proteomics (SILAC) to unravel the mechanism whereby introduction of AGR2 alters cell phenotype, allowing identification of dominant pathways affected by AGR2 signalling. Using pathway analysis tools, the dominant pathway suppressed by wt-AGR2 expression highlighted the p53-signalling axis. DNA damage induced p53 stabilisation and p21 induction by cisplatin treatment confirmed the influence of AGR2 gene expression. Further data analysis identified the outlying protein expression changes identified by SILAC was the anti-viral cell cycle regulator TSG101 (tumour susceptibility gene 101), and confirmed by immunoblotting. Transfection and silencing studies of TSG101 confirmed that TSG101 attenuates p53 function. These data provide a mechanism to explain the most dominant pathways reprogrammed by AGR2 expression, incorporating ER stress response, proliferation markers and p53 pathway attenuation.

Further advances were made in analysis of the function, regulation, and drugability of AGR2 protein. Assays were devised to define the subunit structure of AGR2 as a

dimer unit; subsequent functional studies defined intrinsically disordered motifs that regulate stability of the dimer. A two-site sandwich microtiter assay (^{2S}MTA) was designed to screen for self-peptides and mutations that regulate oligomer stability. These assays were used to identify the first biochemical property of AGR2 being that the dimer unit is required for maximal binding to the AAA+ protein, and well characterised AGR2 interactor, Reptin. In addition, based on this dimeric structure, a novel solution based dimerisation assay was developed to identify natural products that are able to disrupt the dimer suggesting that AGR2 itself can be targeted in principle with small molecules for therapeutic purposes.

Chapter 1: Introduction

1.1 Cancer

1.1.1 Overview of the Molecular Basis of Cancer

The term *cancer* groups a broad spectrum of diseases states, sharing one collective manifestation: the uncontrolled, unregulated growth of cells into a malignant tumour mass. Cancer accounts for 13% of worldwide mortalities (1) and incidence is predicted to rise due to an aging population (1;2). This demonstrates the importance of developing our understanding of oncogenesis through novel research in the fields of biochemistry and biomedicine. The emergence and maintenance of tumours broadly requires the coupling of dysfunctional gene regulation with epigenetic abnormalities (3), resulting in a diverse variety of molecular signatures whose resultant events drive carcinogenesis. The exact mechanism whereby cells acquire these malignant phenotypes is not yet fully understood, yet the heterogeneity of tumour genomes indicate that cells accumulate mutations over time, and the dominant ones conferring a selective advantage are evolutionarily conserved (4). Mutations in pro-oncogenes release the restriction on growth promoting, or anti-apoptotic, genes resulting in oncogenes. Conversely, mutations in tumour-suppressor genes, whose normal function is to constrain cell proliferation within appropriate margins, result in inactivation and loss of function (5-8). Additionally, tumourigenesis can also be linked to changes in gene expression not accompanied by mutations in the DNA blueprint of the cell (9;10), termed epigenetics. Epigenetic mechanisms modulate gene expression by altering the accessibility of DNA to transcription factors; well studied examples include histone modifications (11) and DNA methylation (12;13). Consequently, transcription initiation and gene expression can be misregulated, resulting in disruption of cell homeostasis and aberrant cell growth.

1.1.2 Hallmarks of Cancer

Cell immortality, constitutive chronic proliferation, avoidance of growth suppressive factors and apoptosis, an ability to invade into surrounding tissues and the capacity to activate angiogenesis (6), together with the more recently described, deregulation of cellular energetics (Warburg Effect (14)) and circumvention of the immune system (15) have been defined, and subsequently researched, as key traits central to the definition of cancer (Figure 1-1). These coupled to the compound micro-environment together with normal cells, compliant cancer stem cells and altered extracellular matrix demonstrate the intricacies of the complex web which drive the cancer phenotype (16). Genetic and epigenetic mechanisms give rise to altered protein expression levels, expression of mutant proteins, and deletion of regulatory or housekeeping proteins, and resultantly feed into cancer promoting pathways including growth factor response, metabolic pathways, transcriptional and translational machinery, protein degradation and apoptosis (6;15).

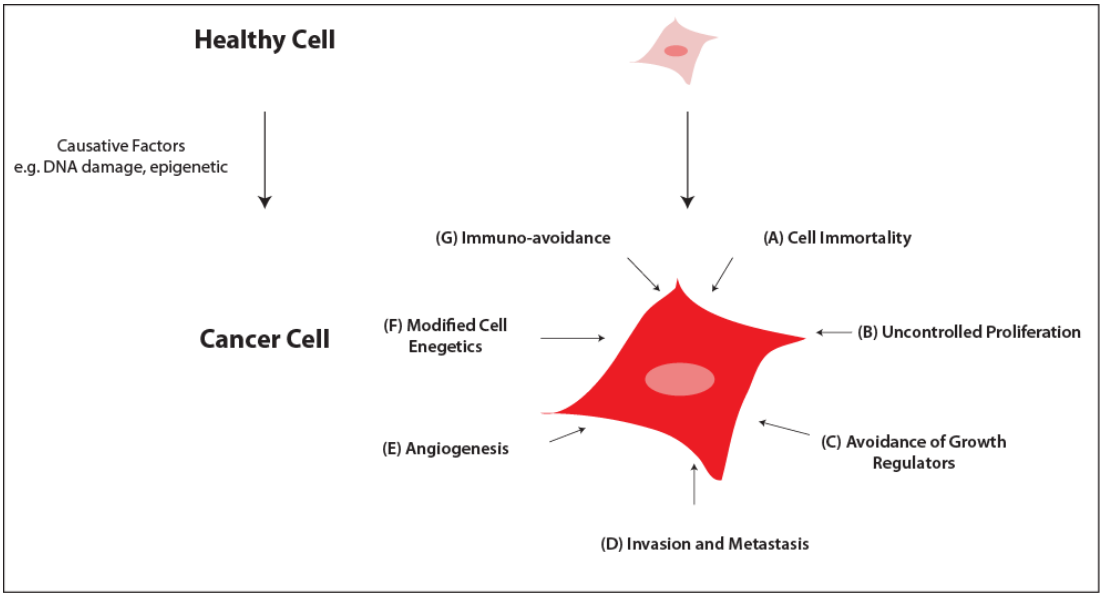


Figure 1-1: An illustration of summarised evolution of a normal health cell into a malignant tumour. Adapted from (6;15).

1.1.3 Cancer Therapeutic Strategies

Despite the disentangling of the key hallmarks of cancer, the heterogeneity and diversity of cancer causing mutations and survival mechanisms of cancer cells presents a difficult challenge for the research and development of novel cancer therapeutics. Developing our understanding of subcellular pathways implicit in cell growth, survival and transformation has provided a rich source of targets for developing strategies for novel cancer therapeutics (17).

1.1.3.1 Broad action, pro-apoptotic chemotherapeutics

As studies develop our knowledge of the molecular basis of tumour progression, biomacromolecular anticancer therapeutics that can target these features have been developed. Paclitaxel is one example of an anti-cancer drug targeting key misregulated pathways prevalent in cancer progression. Paclitaxel targets the microtubule component, tubulin, stabilising the polymeric structure of the cytoskeleton preventing disassembly during mitosis (18). As a result, cells are arrested at the mitotic checkpoint, G_2/M (19), and subsequently prevented from completing cell division and directed toward apoptosis (20). Comparatively, cisplatin is used therapeutically to inhibit cell proliferation by targeting DNA. Cross-linking DNA results in the blocking of DNA replication (inhibiting cell proliferation), and activating DNA repair pathways eliciting apoptotic events (21). Bortezomib instead targets the 26S proteosomal subunit (22), preventing the recycling of misfolded or aberrant protein expression. As a result, cells activate stress response pathways which inevitably lead to apoptosis. These chemotherapies act on the broad understanding that as cancer cells are replicating more than normal cells, they require more cytoskeletal rearrangement, more DNA replication and have in increased rate of protein recycling respectively. Theoretically, driving cancer cells towards an apoptotic pathway allows tumour growth to be controlled. However, since these processes are not only limited to cancer cells, but shared with normal cells also, off-target and side effects would not be unexpected.

1.1.3.2 Drug compounds specifically targeting cancer-associated mutations

Drugs with more specific targets and modes of action have also been developed, these tend to target specific protein cofactors such as kinases, specifically ATP binding sites incorporated into these enzymes, and cell surface receptors mutated in carcinogenesis. The discovery of the mutation of functions regulating protein kinase activity in cancers (23) has emerged as one of the mechanisms whereby cancer cells circumvent normal physiological control, making it a suitable specific target for anti-cancer therapeutics. Mutated kinases provide a selective advantage to cancer cells compared to normal cells resulting in an 'oncogene addiction' phenomenon (24), where cells are reliant on a few genes for the maintenance of the malignant phenotype. A mutation, such as the DNA translocation which forms the BCR-Abl fusion protein, allow the encoding of a mutated, constitutively active tyrosine kinase which drives unregulated cell division through the activation of a number of cell cycle controlling proteins and enzymes (25). Imantinib is an example of a drug developed specifically targeting this single aberrant tyrosine kinase, inhibiting phosphorylation of substrate proteins necessary for the cell signalling cascade that drives cell proliferation (26). Several other examples of protein kinase mutations in cancer are evident including epidermal growth factor (EGF) receptors (27) and BRAF (28). Targeted therapeutics can also specifically bind cell surface proteins on mutated in cancer. Trastuzumab, a monoclonal antibody therapy which interferes with the overexpressed HER2/neu cell surface receptor (29), constricts the over-activation of EGF receptors and downstream activation of proliferative pathways, arresting the cell in G_1 phase of the cell cycle. These drugs target cancer cell specific targets, therefore demonstrate an increased affinity for aberrant cells, compared to the more broadly acting drugs (30). However, due to the diversity of cancer, these drugs are only useful in disease exhibiting a specific mutation limiting their therapeutic potential to particular classes of cancer manifestation.

A key objective of cancer research is to develop our understanding of novel drug targets, to broaden the range of therapeutic interventions available, particularly when cancers show resistance (31) or do not respond to current options.

1.1.3.3 Protein-protein interactions and hub proteins as emerging cancer therapeutic targets

The complex webs of protein networks that fuse together to form the molecular basis of all cellular functions are presenting a promising emerging basis for rational anti-cancer drug design (32). The connectivity between these molecular components and dysfunction therein, translates into the global expression of disease. An interactome describes all of the molecular interactions regulated by a specific protein of interest (33;34), highlighting that some central 'hub' proteins interact with hundreds of partners to form the highly connected axes of cell signal transduction. Small manipulations, or mutations, in the promiscuity or expression of these hub proteins can have a significant effect on the normal homeostatic mechanisms of a healthy cell. A mutant variant of a hub protein could be considered to rewire the signal transduction mechanism of a cell, through the protein-protein interaction landscape, contributing to cancer-promoting activity. Indeed, genes identified which are associated with, and responsible for, tumourigenesis have a high likelihood of encoding hub proteins (32).

The tumour suppressor protein, p53, has been described as a key hub protein, mediating gene expression, cell cycle regulation and regulation of apoptosis and senescence (35). p53 interacts with in excess of 300 interaction partners (36), and is the most frequently mutated or deactivated in one in every two human cancer incidences (37). This provided the basis for the development of the p53-reactivating drug PRIMA-1 (p53 re-activation and induction of massive apoptosis-1) (38), which selectively covalently inhibits the growth promoting, mutated p53 gene product, promoting apoptosis and death of the aberrant cell.

Molecular chaperones form another target for anti-cancer drug compounds utilising protein-protein interaction networks. Chaperone proteins are essential in the maturation of nascent peptide chains to form proteins, and in the response of cells to stressful stimuli in order to maintain cellular homeostasis and promotion of cell survival (39). The heat-shock protein 90 (Hsp90) is widely reported to be a hub protein involved in cancer related protein folding pathways including p53, BRAF and AKT (35;40). Cancer cells are regularly exposed to proteotoxic stress, through

the increase in protein synthesis and accumulation of unfolded or misfolded proteins, therefore they become dependent on the role of chaperone proteins for proliferation and division of cells (41;42). Accordingly, inhibitors of Hsp90 have been developed, geldanamycin and 17-AAG (17-allylamino-17-demethoxygeldanamycin) (43), which target the ATP-binding pocket of Hsp90, blocking ATPase activity and controlling proteostasis and allowing the curbing cell growth (44-46).

1.2 p53

p53, named after migrating 53 kDa on a SDS-PAGE gel despite a molecular mass of 43.7 kDa, was first published in 1982 as a protein forming physical interactions with transforming protein SV40-T-antigen of a DNA tumour virus (47), allowing perturbation of normal cell signalling pathways and promoting replication of viral DNA. Subsequent intensive research has identified p53-mutation (48) or deletion (49) as a central event in the predisposition of 50% incidences of cancer. However, some cancers never select for p53 mutations, and yet exhibit a functionally inactive p53 protein (50). These carcinomas utilise mutation independent mechanisms. The common theme is the sequestration of wild-type p53 protein away from the nucleus resulting in abrogation of normal suppressive action (51-54).

Molecular characterisation studies of p53 protein have begun to disentangle the structural motifs or p53 which modulate its described functions.

1.2.1 p53 structure

Despite a high degree of disorder in the predicted structure of p53, the 393 amino acid residue protein presents several modular domains giving insight into its activity (Figure 1-2A and 1-2B). The (i) N-terminal region incorporates two transactivation domains required for interaction with transcriptional machinery such as p300 (55), and also provides a binding site for the p53-regulating protein, mouse double minute

2 (MDM2) (56). Additionally the (ii) core DNA binding domain of p53 incorporates several conserved domains (Box II, III, IV and V) (57;58) involved in direct DNA contact, and is the site of most tumourigenic mutations (48) (Figure 1-2C). The (iii) tetramerisation domain of p53 describes the α-helical motif necessary for the oligomerisation of p53 monomers into the collective transcriptionally active functional unit (59-63), and can also function as a protein docking site regulating protein-protein interactions (64-66). Finally, the C-terminal (iv) regulatory domain forms a further protein docking site for regulation of p53 activity (67;68)

The high level of disorder has resulted in the core domain and tetramerisation domain of p53 being structurally characterised individually (69;70). It is important to note, that significant effort has been directed at the post-translational modifications of these modules affecting activity of the protein, including ubiquitination and ubiquitin-like modification (neddylation, sumolation), methylation, acetylation and, most significantly, phosphorylation. Phosphorylation and acetylation of the C-terminal domain allosterically stimulate DNA binding to the core domain (71;72), whereas N-terminal post-translational modifications fine tune p53 activity, often in response to stress (73).

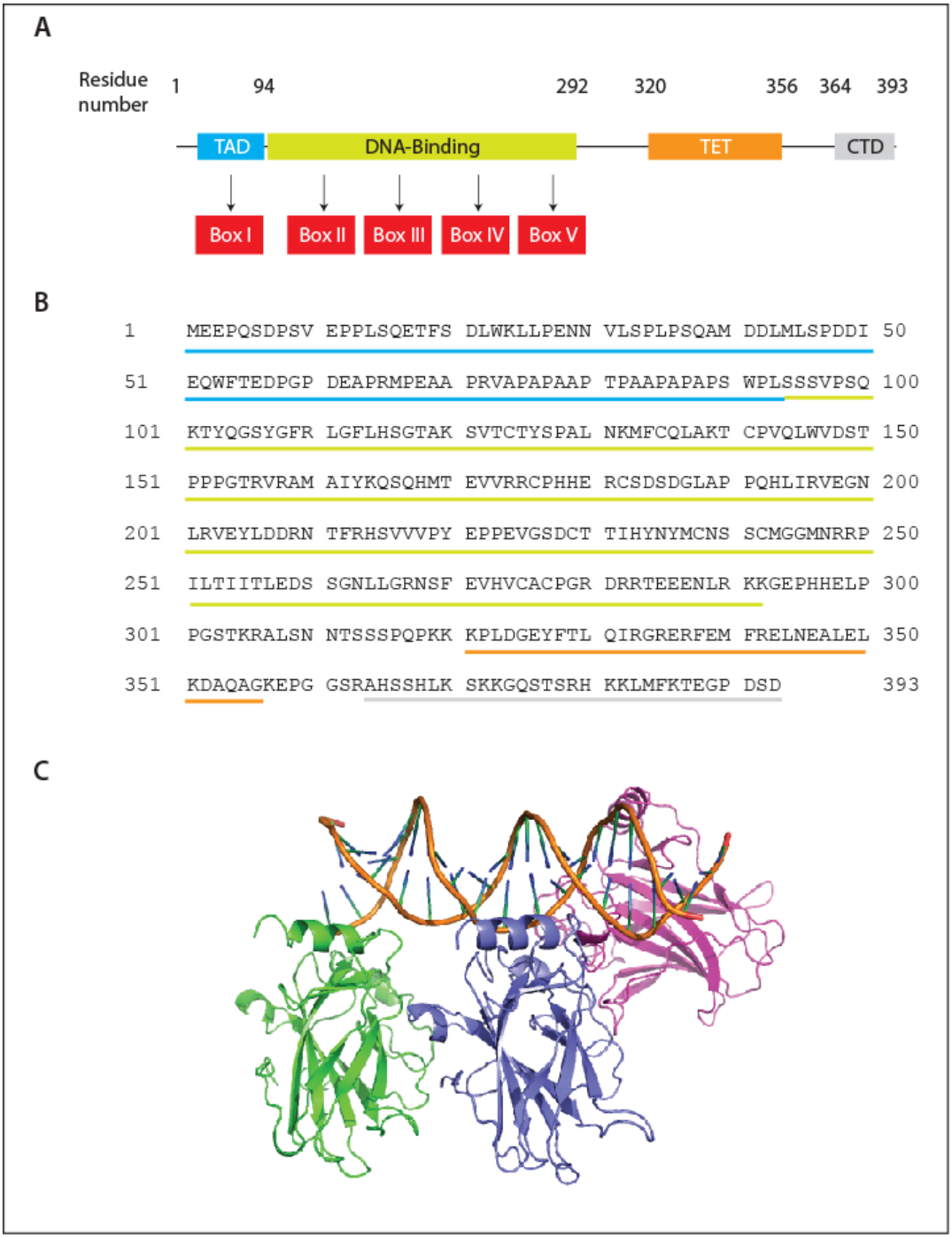


Figure 1-2 p53 protein structure. (A) Illustration of the modular structure of p53 domain organisation highlighting the transactivation domain (TAD, blue), the DNA-binding domain (green), the tetramerisation domain (TET, orange) and the C-terminal domain (CTD, grey). (B) The primary sequence of human p53, with domains coloured as above. (C) Cartoon model of three core DNA binding domains of p53 interacting with DNA (PDB code: 1TUP (74)).

1.2.2 p53 function

1.2.2.1 p53 as a hub protein

All proteins interact with other molecules in order to carry out their perceived role, however, a small number of proteins, hub proteins, form relationships with hundreds of interacting partners resulting in a focal point of the interaction network (75). This promiscuity indicates the complexity of the function of these hub proteins. p53 is described as an example of a hub since it interconnects numerous signalling pathways signifying the diversity of functions with which p53 has been implicated. The interactome of p53, that is all of the published protein-protein interactions, now exceeds 300 partners (36) providing a broad database within which p53 signalling can be interrogated. It is evident from such a broad range of interaction partners that slight modification in the gene, the expression of the gene product or the subcellular localisation of p53 protein can have significant downstream consequences for the cell.

1.2.2.2 p53 tumour suppression function

Mammalian cells have evolved mechanisms to tightly control the maturation of cells during important functions such as regeneration, replenishment and growth. Concurrently, inaccuracies in gene expression arising from DNA replication errors, metabolic activities and environmental factors, are remedied through complex, controlled cellular repair systems. If significant DNA damage has occurred, cells can enter two possible states: (i) an irreversible, non-dividing arrested state, termed senescence, or (ii) cells can commit suicide through a process known as apoptosis (a third process of autophagy, or 'self-eating', the processing of bulk degradation using the lysosomal machinery, is reportedly also thought to play some role (76)). Due to the reported prevalence of mutation of the p53 allele of human cancer biopsies (48), studies have been undertaken to decipher the growth controlling properties of this master inhibitor of tumour development (77;78). Under normal circumstances, p53 is a protein with an extremely short half-life (79), its expression is tightly controlled and excess protein rapidly sequestered. Surfeit protein is recycled by targeting for

ubiquitination by the ring-finger ubiquitin E3 ligase MDM2, and trafficked out of the nucleus to be degraded by the 26S proteosome (80) (Figure 1-3A).

In response to DNA damage by ultraviolet light (81) and γ-radiation (82), p53 forms a central node whose expression is stabilised by post-translational modifications, predominantly kinase and acetyltransferase activity which phosphorylate and acetylate respectively, and this stabilisation mediates cell cycle arrest at G_1 (82). p53 tetramerises and accumulates in the nucleus (83), leading to subsequent transcriptional activation of p53 target genes (Figure 1-3B). Growth arrest is principally induced via the best characterised of p53-dependent downstream targets, p21-WAF1 (84). Elevated levels of p21 inhibit the cyclin E/cdk2 and cyclin A/cdk2 kinases preventing the phosphorylation of client proteins promoting cell cycle progression (77;85). Of the approximately 150 gene targets of p53, most are associated with cell cycle arrest, apoptosis, and DNA repair, which maintain the integrity of the cell (86) (Figure 1-3B). Key transcriptional targets include the Bcl-2 proteins p53-upregulated modulator of apoptosis (PUMA) (87), NOXA (88), BID (89) and BAX (90), while suppressing the anti-apoptotic genes Bcl-x and survivin (91;92), along with other members of the pro-apoptotic machinery. Bcl-2 proteins perform their pro-apoptotic function by initiating processes such as caspase activation and mitochondrial outer membrane permeabilisation (MOMP) (93;94), whereby key structural components of the cell are disrupted.

Apoptosis is firmly established as the central mechanism of control of damaged cells preventing aberrant cell growth. However, none of these gene targets can fully explain the pro-apoptotic effects of p53 (91) indicating that p53 acts via multiple pro-apoptotic pathways. Intriguingly, apoptosis can also be mediated independently from transcription (52;95-99). Cytosolic and mitochondrial localised p53 interacts with members of the Bcl-2 family, BAK, BAD and BAX (100) to modulate the activation of apoptosis. This response protects the cell from uncontrolled proliferation and neoplastic transformation following genotoxic stress.

Further, p53 can also prevent tumour development independently from apoptosis. PUMA knockout mice do not exhibit uncontrolled proliferative growth (101),

indicating that the p53 anti-tumour response is more complex than promoting defective cell death. p21-WAF1 (84), GADD45 and 14-3-3σ proteins, induced by p53, can activate a reversible cell cycle arrest (82;84;102) to pause cell growth to allow repair. Subsequently, senescence is triggered if the repair is not successful (103).

p53 also limits proliferation through the inhibition of new blood vessel formation through upregulation of anti-angiogenesis factors vascular endothelial growth factor-A (VEGF-A) and hypoxia inducible factor 1α (HIF-1) (104-106). Thereby, p53 can influence the distribution of nutrients and oxygen in a controlled manner.

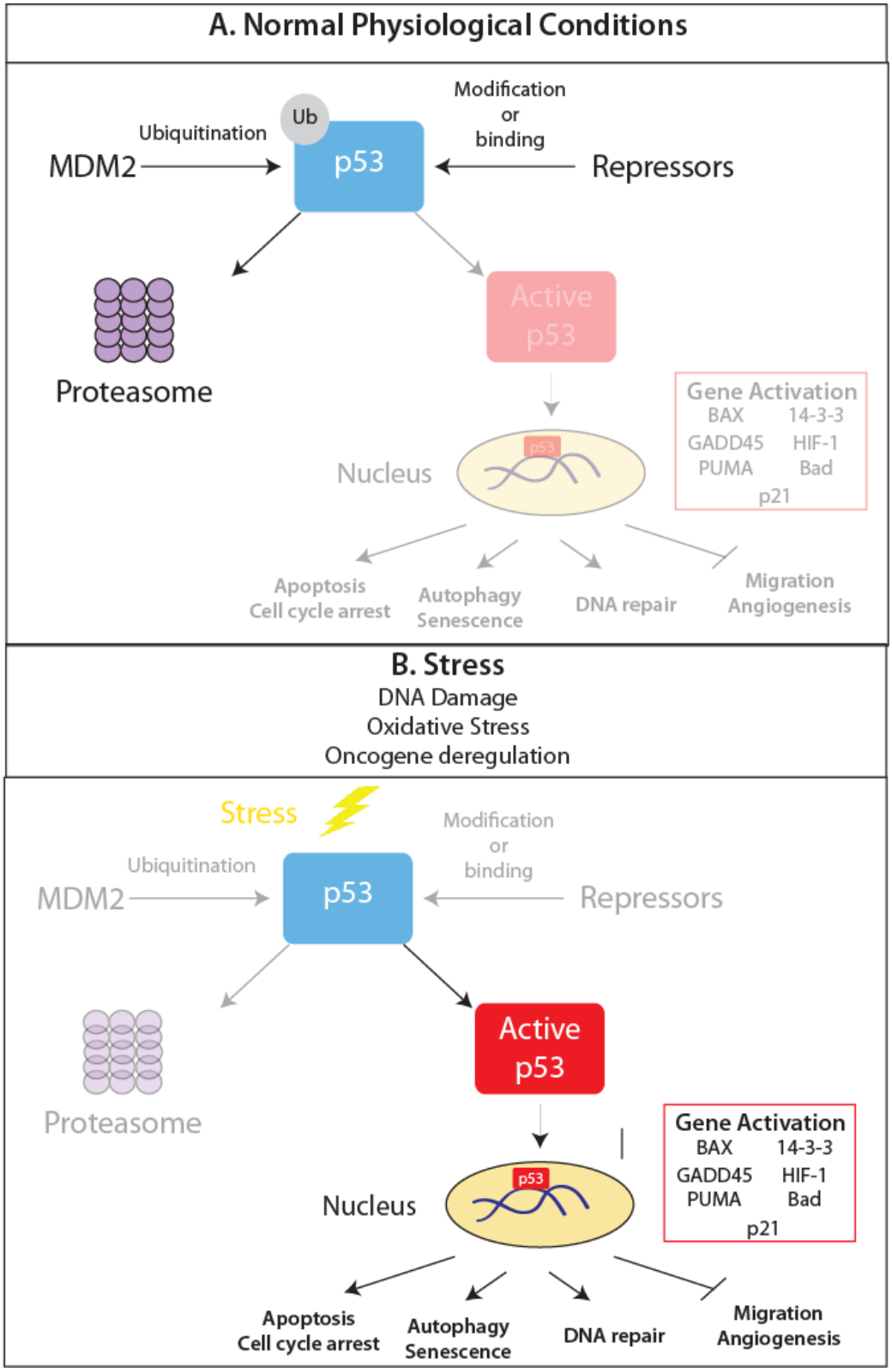

A. Normal Physiological Conditions
Ubiquitination
Ub
MDM2
p53
Modification
or
binding
Repressors
Proteasome
Active
p53
Gene Activation
BAX
14-3-3
GADD45
HIF-1
PUMA
Bad
p21
Nucleus
p53
Apoptosis
Cell cycle arrest
Autophagy
Senescence
DNA repair
Migration
Angiogenesis
B. Stress
DNA Damage
Oxidative Stress
Oncogene deregulation
Stress
Ubiquitination
MDM2
p53
Modification
or
binding
Repressors
Proteasome
Active
p53
Gene Activation
BAX
14-3-3
GADD45
HIF-1
PUMA
Bad
p21
Nucleus
p53
Apoptosis
Cell cycle arrest
Autophagy
Senescence
DNA repair
Migration
Angiogenesis

Figure 1-3 The published roles of p53 under normal physiological conditions and following stress-inducing conditions. (A) Under normal, undamaged cellular conditions, p53 is maintained at a low level of expression. Gene expression constitutively suppressed and surfeit protein is labelled for proteosomal degradation through ubiquitination by MDM2. (B) In contrast under stressful conditions, p53 is transcriptionally activated and protein turnover is suppressed allowing enrichment of p53 protein in the cell. This active p53 translocates to the nucleus, allowing interaction with DNA and upregulates the expression of gene targets involved in several cellular repair pathways.

1.2.2.3 p21 involvement in p53 signalling

p21 was identified as a p53-responsive target in a screen to identify p53 target genes, subsequently described as WAF1 (wild-type p53 activated fragment 1) (84) and is a multifunctional protein implicit in cell cycle arrest, apoptosis and promotion of differentiation and cellular senescence. However, its role is most defined in the p53 pathway where p21 is a direct transcriptional target of p53, and is strongly induced by DNA damage in cells expressing wild-type p53 (107). It has been characterised as a cyclin dependent kinase inhibitor, induced by p53, and shown to mediate p53-dependent cell cycle arrest at the G_1/S checkpoint by inhibition of the activity of cyclin/CDK2 complexes (108). The p21 promoter contains five p53-binding sites at positions -4001, -3764, -2311, -2276 and -1391 base pairs (109) and at least one of these sites is required for p53-dependent induction after DNA damage (110;111). p21 protein levels are regulated at the post-translational level by proteosome mediated degradation by ubiquitin-dependent and –independent means (112;113). p21 is also thought to have a regulatory role in the nuclear localisation of p53, driving a positive feedback loop of p53 expression (114).

1.2.2.4 Misregulation of p53 in cancer

Inactivation of p53 can occur by genetic deletion, or down regulation of the expressed protein, or more dominantly, in the mutation of the genomic p53 blueprint, resulting in a protein with aberrant functions (48). p53 mutants are widespread in cancer, and hot spots of mutation have been identified, and map in particular to regions of the core DNA binding domain (35). Mutant alleles of p53 are often dominant over the wild-type, and in these cases the gene product with aberrant function may act as an oncoprotein (36). Since these mis-sense mutants offer a survival advantage, such that mutant p53 inactivates normal sequence specific DNA binding (‘transcriptionally-inactive’), promoting cell proliferation and suppressing the homeostatic function of wild-type p53, these are commonly selected for. Studies characterising mutant p53 alleles have provided key insights of the roles of mutant p53, and inversely wild-type p53, in the field (36); including (i) unfolded, mutant p53 accumulation in the nucleus of cancer cells *in vivo* (115), (ii) mutant p53

dominates wild-type p53, by forcing unfolding of the wild-type protein and the formation of hetero-tetramers (116), (iii) mutants lack sequence specific DNA-binding activity (117), and (iv) mutant p53 exhibits a distinct interactomic landscape compared to the wild-type (118;119).

However, some cancers maintain the wild-type p53 allele, or mutation occurs late in carcinogenesis such as colorectal cancer (120) and glioblastoma multiforme (121), yet have disrupted p53-dependent function. This provides an indication that a secondary mechanism exists to ablate p53 function. Under normal conditions, wild-type p53 is predominantly kept inactive in the vast majority of human cells, however a variety of stresses induce p53 activation and resultant p53-dependent activity. Cell stresses such as DNA damage, oncogene activation and hypoxia have been uncovered as regulators of p53. These regulators are not thought to heavily influence p53 at the transcriptional level, but rather modulate the level of available p53 protein by changing the half-life of the protein (122). The best understood, and most important, degradation pathway of p53 occurs via the ubiquitin-proteasome system, in which p53 is a substrate for the E3 ubiquitin ligase, MDM2 (80). E3 ligases function in an enzymatic cascade, with an E1 and E2 enzyme, which catalyse the transfer and covalent linkage of a ubiquitin molecule to lysine residues of the substrate protein, signalling for translocation to the proteosome and subsequent degradation (123). MDM2 also contains a signalling sequence similar to the nuclear export signal of various viral proteins (124) and after binding to p53, signals for its nuclear export (125;126). Therefore, the availability and activity of MDM2 is an important factor in the normal control of p53 activity, maintaining the tight regulation in healthy cells under normal conditions. The mouse model containing a gene knockout mutant of MDM2 describes an embryonic lethal phenotype, which can be rescued by subsequently knocking out of p53 allowing live births (127). The transcription of MDM2 is controlled in an auto-regulatory feedback loop by p53, which acts to upregulate MDM2, thus p53 expression acts to negatively regulate itself via an autoregulatory feedback mechanism (80). This interplay influences a vast array of molecular pathways in humans including autophagy, DNA damage response, differentiation, senescence, cell-cell interactions, apoptosis and metabolism

(36). Studies have confirmed that the over expression of MDM2 *in vitro* acts to block p53-mediated cell cycle arrest and apoptosis (128). This is reflected in clinical studies of tumours, predominantly soft-tissue tumours, osteosarcomas and oesophageal carcinomas among others, where MDM2 acts to suppress wild type p53 activity, and is associated with increased cell proliferation and tumour volume. (129-131).

1.2.2.5 Therapeutic targeting of p53

Based on these understandings of p53 as a tumour suppressor and its misregulation in tumours, there are two key ways to target p53 for therapeutic treatments of cancer: (i) the inhibition of mutant p53 and (ii) reactivation of wild-type p53. Specifically targeting mutant p53 allows the binding and silencing of oncogenic function, a methodology which has been successfully utilised in the development of the drug CP-31398 (Pfizer) (132-134), enabling the activation of p53-dependent transcription, apoptosis, cell cycle arrest and reduction in tumour growth in mice. Mutant p53 can also be targeted by structure based drug design and *in silico* modelling, specifically targeting individual mutants such as the Y220C p53 mutant binding pocket to reactivate p53-dependent activity (135). The cysteine substitution forms a pocket which does not affect the overall structure, but demonstrates reduced functional activity (35), and the development of the molecule PhiKan059 can bind and inactivate p53-Y220C (135;136). Actively targeting amino acid residues deemed essential for function has yielded the mutant-p53 targeted therapeutic PRIMA-1 (38). PRIMA-1 modifies reactive cysteines to reactivate p53 function. This strategy is hugely resource consuming due to the breadth of different cell types, microenvironments and mutations that exist. Variations in selective pressures driving the selection of mutation may be essential in the understanding of targeting mutant p53.

In wild-type p53 expressing cancers, many strategies focus on MDM2, to reactivate p53 transactivation, cellular localisation and mRNA translation (137;138). Nutlins were the first small molecule drugs targeting the MDM2-p53 interaction, allowing

dissociation of components (139;140) and subsequent inhibition of tumour growth (141;142). A small molecule screen of novel p53 reactivators identified RITA (reactivation of p53 and induction of tumour cell apoptosis)(143-145), initially thought to disrupt the MDM2-p53 interaction, however, RITA also induced DNA damage (143;144) suggesting the mode of action of RITA is more complex than initially described. Importantly, RITA could downregulate p53 inhibitors MDM2, and WIP1, resulting in p53-mediated cell death (146). However, due to the vast complex interactome of MDM2 (36), some of which functions independently of p53, off target effects may be common. In fact, some studies have suggested that small molecule-induced interaction with MDM2 may in fact fuel metastasis by interfering with the regulation of other MDM2 substrates (e.g. Slug) (147). The degradation of Slug, an invasion promoter, is mediated by MDM2-controlled ubiquitination and proteosomal targeting. Thus, silencing of MDM2 allows stabilisation of Slug, and subsequent promotion of cancer cell invasiveness (147).

There are obvious difficulties in understanding the possible therapeutic strategies of p53 reactivation in cancers. It is therefore imperative to develop our understanding of the protein landscape surrounding p53, through mutations, interactors and regulatory proteins. Analysing how these correlate to p53-dependent activity, normal cellular function and cell homeostasis and whether these compounds might indirectly be useful of therapeutic strategies to reactive p53 function.

1.3 Anterior Gradient-2

Broad research approaches into possible roles of anterior gradient-2 protein (AGR2) have been undertaken since its discovery as a developmental factor in 1989 (148) and consequential learning of its role in the specification of cement gland and forebrain in developing embryos (149). Subsequent studies have implicated AGR2 in a range of pathways and pathologies described herein. Specifically, and most relevantly, AGR2 was identified as a negative regulator of p53 (150), driving approaches to

understand it as a potential chemotherapeutic target for the reactivation of p53-dependent activity in cancer. Firstly, I present a summary of the discovery and evolution of AGR2, and related proteins, before analysis of studies of AGR2 roles and implications in human diseases.

1.3.1 AGR2 in development and limb regeneration

AGR2 belongs to the wider family of anterior gradient proteins, containing several members including the ancestral member XAG-2 (Figure 1-4). The anterior gradient genes, XAG-1, XAG-2 and XAG-3, were first discovered in *Xenopus laevis*, and named according to their specific expression patterns during early development (148;151;152). XAG-2 was further validated by efforts to characterise novel genes required for the development of *X. laevis* embryos (149). The gene was described as being specifically synthesised in the cement gland of developing embryos during the late gastrula stages, encoding a secreted protein whose transcriptional levels were upregulated in response to organiser-secreted molecules chordin, noggin, follistatin and cerebrus gene products. As such, the protein was incorporated into a pathway involved in ectodermal patterning of *X. laevis* embryos defining a role of XAG-2 in specification of dorsoanterior ectodermal fate. XAG-2 expression is a required determinant in the definition of the *Xenopus* ectoderm to form structures including the forebrain and the cement gland (149). A homologous gene to XAG-2, namely *XAgr2*, was identified as strongly expressed in the *Xenopus* cement gland, and additionally in the otic vesicles which later develops to the cochlea, and notochord, particularly in the tailbud and neurula stages (153). This gene was found to have greater homology to human and mouse AGR2 than XAG-2 (153).

Anterior gradient proteins were brought to the attention of the wider developmental community when the *Notophtalmus viridescens* (newt) homologue, nAG, was identified as the first ever cofactor with the ability to promote limb regeneration following amputation (154). Newt limbs regenerate from dedifferentiated fibroblast, muscle, skeletal and Schwann cells into an earlier stem cell form, yet cofactors influencing re-differentiation are poorly understood. Post limb amputation, severed

axons retract inward from the wound then grow back along the new nerve sheath. This nerve sheath is comprised of nAG-expressing Schwann cells as the blastema, the regenerating pleuripotent cells, divide. nAG was discovered as a secreted ligand for the surface GPI-anchor protein Prod-1 (by yeast-2-hybrid studies) (154), where Prod-1 is a critical determinant of the proximodistal identity of regenerating limb structures by mediating cell-to-cell interaction of tissues (155). nAG is highly expressed in the wound epidermis at the surface of the amputated/regenerating limb. When cells isolated from the blastema are cultured *in vitro*, nAG protein acts as a mitogen to promote cell proliferation. Further, when nAG is artificially expressed in the regenerating limb which has been denervated, removing the Schwann cells, this results in the re-establishment of missing glands, and strikingly, the regeneration of the proximal-distal axis and the formation of digits. This indicates that nAG, secreted by nerve cells of the regenerating limb, specifically acts to differentiate limb shape and structure (154). However, for completeness, studies should be undertaken to silence nAG expression *in vivo* in the newt limb following amputation using gene editing in situ with TALENs (transcription activator-like effector nuclease) (156). This recently published technique would not have been available to researchers at the time of publication. The role of anterior gradient proteins in limb regeneration has since been widened to other amphibians as well as newts, and has been described as having a universal role in body appendage regeneration in amphibians (157).

1.3.2 AGR2 homologues in other vertebrates

The expression of homologues of XAG-2 in further vertebrates have been described (Figure 1-4). In zebrafish, *Dario renio,* zAG-2 was identified in most organs containing secretory cells, including epidermis, olfactory bulbs, otic vesicles, pharynx, oesophagus, pneumatic duct, swim bladder, and intestine (158), and within the stomach is required for the terminal differentiation of goblet cells of zebrafish embryos (159). mRNA transcripts of two AGR2 homologues were observed as differentially over-expressed in healthy Atlantic salmon (*Salmo salar*) and salmon which were affected by amoebic gill disease (160). These mRNA were cloned and

sequenced, and shown to be predominantly expressed in the gill, intestine and brain of healthy fish. In the gibel carp, *Carassius auratus gibelio*, the AGR2 homologue was shown to be expressed in the mucus secreting hatching gland, and later the pharynx, swim bladder and pronephric duct. In the mature carp, AGR2 expression was primarily localised in the intestine, ovary and gills (161).

In mice, initial transcript analysis revealed expression of the AGR2 homologue (mAGR2, GOB4) in the intestines, colon and stomach of the fully grown mouse (162). Murine models of normal mice, germline and inducible AGR2 knockout mice, along with *in vitro* cell models have begun to describe the role of mAGR2 in normal tissue. AGR2 expression is localised to distinct cells of the highly folded epithelial lining of the small intestine. Several cell types compose the small intestine including nutrient absorbing enterocytes, and cells with key secretory cell lineages of the goblet, Paneth and enteroendocrine (163) (also described in zebrafish goblet cells (159)) . Murine AGR2 is strongly expressed in these differentiated goblet, Paneth and enteroendocrine cells, as well as intestinal progenitor and stem cells (164). Gene silencing studies, utilising AGR2-specific siRNA knockdown have described AGR2 as performing a role in the maintenance of intestinal homestasis, with respect in particular to goblet and Paneth cells (165;166). AGR2 depletion, results in the loss of the goblet cells function to secrete intestinal mucus (165), and significantly changes the morphology of these cells (165;166). AGR2 was immunostained residing in the ER of these secretory epithelial cells and is essential for the *in vivo* production of the intestinal mucin, MUC2 (165). The study by Zhao *et al.* (166) describes additional aberrations causing symptoms in the small intestine of AGR2 knockout mice, the Paneth cells morphology was drastically altered, and severe intestinal inflammation reported. Specifically, the Paneth cells display atypical localisation, with differentiation spanning the entire length of the small intestine. Subsequently, *agr2 -/-* mice displayed severe inflammation of the small intestine, consistent with the role of Paneth cells in the inflammatory response (167;168). Furthermore, loss of AGR2 expression results in disruption of enterocyte homestasis, reducing proliferation, increasing apoptosis and blunting of villi (166). Furthermore, elevated endoplasmic reticulum stress response is present in cells lacking AGR2, in

some intestinal cell types. Interestingly, phenotypically, AGR2 knockout mice present symptoms resembling human Crohn's disease, suggesting a role of AGR2 in the aetiology of inflammatory bowel disease (discussed in 1.3.5.3).

These studies present convincing data, indicating that AGR2 and its homologues are not exclusively secretory proteins, but in higher vertebrates function intracellularly, in this case in the proper maturation of proteins in the ER.

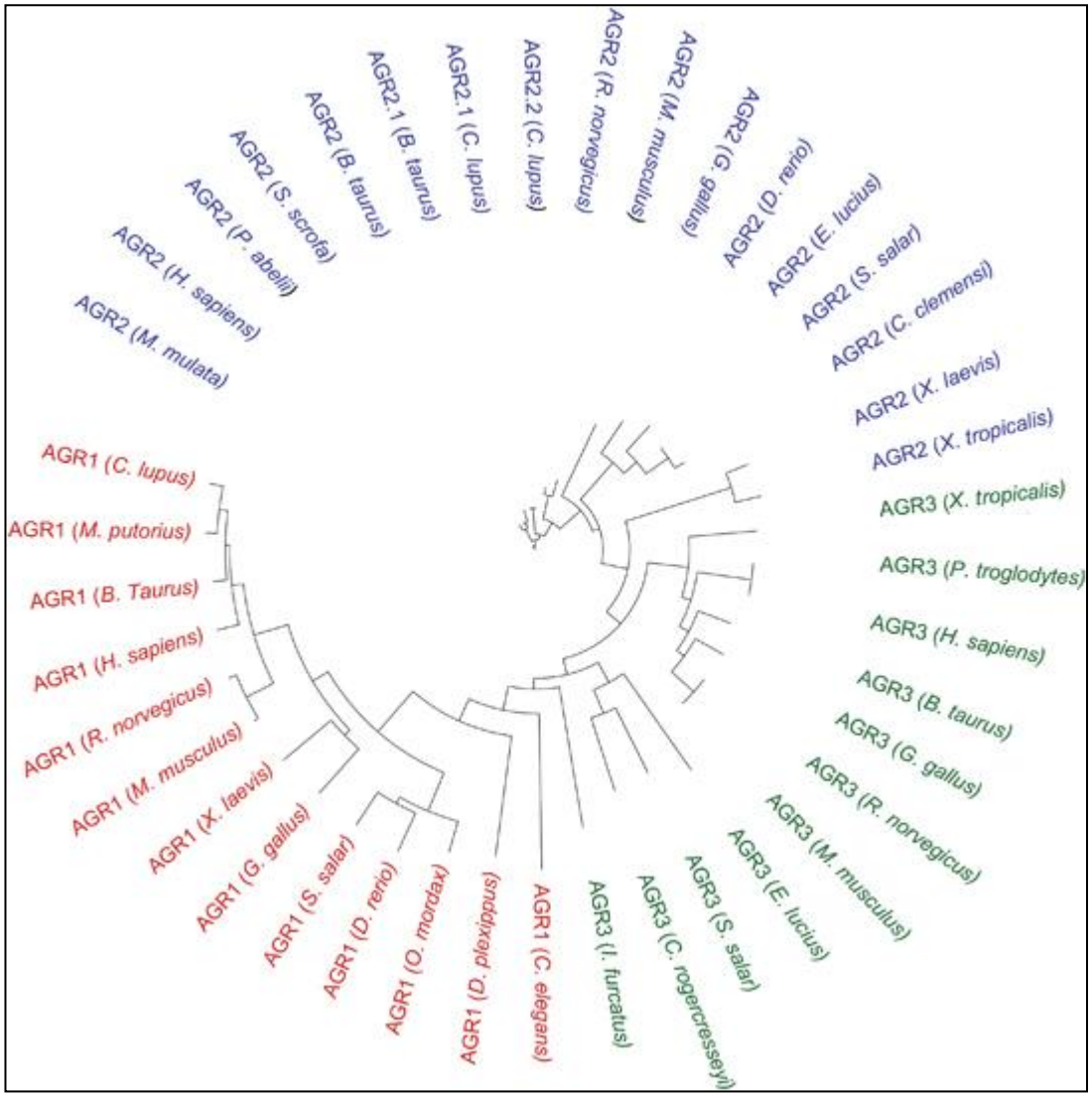


Figure 1-4 Phylogenetic tree of the Anterior Gradient family of proteins adapted from Chevet *et al* (169). ClustalW alignment of the 41 related proteins deduced amino acid sequences derived from the public databases. Colours represent which category each homologue relates to (blue: AGR2, green: AGR3 and red: ERP18/AGR1). Figure created using the Interactive Tree of Life (iTOL) web server (http://itol.embl.de/).

1.3.3 Human AGR2

1.3.3.1 Gene architecture and regulation

The *agr2* gene emerges during the evolution of the chordates, with *agr3*, from the founder gene ERP18 (AGR1), and is confined to vertebrates. There has been no expansion of *agr2* and *agr3* genes in vertebrates, indicating that strict selection pressures exist in maintaining essential functions during evolution by natural selection. The human *agr2* gene spans 50 kb of genomic DNA, located to chromosome 7p21.3 (170), and codes for two mRNA isoforms containing eight exons and seven introns, and both isoforms contain the same open reading frame (169). An annotated model of *agr2* exhibits an additional 5' exon, confirmed by RT-PCR in a tissue panel (171). The genes for *agr2* and *agr3* lie in close proximity on chromosome 7, and share 71% similarity, yet show little redundancy (172). A luciferase reporter gene construct driven by the *agr2* promoter assay indicated that the forkhead box transcription factors FOXA1 and FOXA2 increased luciferase activity thus may regulate the transcription of AGR2 mRNA (171). In addition, the *agr2* promoter region also exhibits a binding site for hepatic nuclear factor 1, related to the FOXA1 and FOXA2 regulators (171). Further chromatin immunoprecipitation (ChIP) studies identified binding sites for the oncogene ZNF217, NF-κB and SOX9 within the *agr2* promoter region (169;173). The increased transcription of *agr2* in response to oestrodiol, and subsequently with the anti-oestrogen drug tamoxifen (174;175), together with its co-expression with oestrogen receptor molecules in breast cancer cell lines (176), and induction by androgens (177) complete the known molecular regulators of *agr2* transcription. Additionally, physiological stresses of serum starvation, coupled with oxygen restriction induced *agr2* transcription five-fold (178). Conversely, this study also interrogated specific inhibitors to known transcriptional regulatory pathways ERK1/2, JNK, p38 and PI3K and only ERK1/2 inhibition resulted in a reduction of *agr2* mRNA expression (178). Tunicamycin (an inhibitor of N-linked glycosylation), DTT (reducing agent) and thapsigargin (a SERCA, sarco/ER Ca^{2+} ATPase inhibitor), all promoters of ER stress, were uncovered as inducers of *agr2* transcription (177;179). Recently, *agr2* has also been described as tumour growth factor-β (TGF-β) responsive in pancreatic cells, whose expression can be downregulated by SMAD4 induction (180).

1.3.3.2 AGR2 protein structure

The *agr2* gene product translates to a 175 amino acid residue protein, with a predicted molecular mass of 19,979.2 Da and a predicted pI of 9.03 (ExPASy Compute pI/Mw. http://web.expasy.org/compute_pi/). A sequence alignment of AGR2 reveals 54% identity and 71% similarity to XAG-2, and 91% identity and 96% similarity to murine AGR2 (181). The primary structure and recent research into the structural motifs of AGR2 highlight several important characteristics of the expressed protein (Figure 1-5). The N-terminal 45 amino acid residues exhibit high levels of predicted disorder (DisoPred2, http://bioinf.cs.ucl.ac.uk/index.php?id=806) (182) (Figure 1-5A). Regions of intrinsic disorder are particularly enriched in proteins involved in protein-protein interactions, cell signalling and transcription (183). Further bioinformatical analysis of the primary protein structure identifies a 20 amino acid residue hydrophobic endoplasmic reticulum leader sequence incorporating a signal peptidase cleavage site at ala-20/lys-21 (184;185) (SignalP 4.1 Server, http://www.cbs.dtu.dk/services/SignalP/) (Figure 1-5B). This leader sequence motif targets nascent polypeptides directly from the ribosome to the ER membrane, via a signal recognition particle (SRP) before co-translational cleavage of the motif by a signal peptidase, allowing protein access to the ER lumen. This process is thought to be essential for the correct maturation of the nascent polypeptide into mature folded protein (186). Biochemical studies transfecting the unprocessed (full-length AGR2) and the signal peptide cleaved mature protein have suggested distinct subcellular distribution (187). This would not be unsurprising, as transfected protein lacking the N-terminal leader sequence would not be directed to the ER lumen, and instead would demonstrate mislocalisation to the cytosol, nucleus and/or mitochondria.

Despite reported secretion during development of lower vertebrates, human AGR2 is predicted as exhibiting and ER localisation, due in part to the N-terminal leader sequence, but complemented by a C-terminal putative ER-retention tetra-residue sequence, KTEL (Figure 1-5B). This motif is conserved from *Xenopus* to humans so assumed to be essential for function (188). The classical canonical H/KDEL motif has been well studied in higher vertebrates (189), yet divergence is not uncommon.

Subsequent studies have compiled a Prosite motif (ExPASy, http://prosite.expasy.org/) of (K/R/H/Q/S/A)(D/E/N/Q)EL, which describes the possible carboxy-terminus peptide sequence that results in ER residence (188). Early studies determined that the H/KDEL motif is recognised by receptor proteins resident in the Golgi complex, which interacts with the client protein and triggers retrograde transport back to the ER lumen (190). Controversially, despite H/KDEL being well described as an ER trafficking motif, a few examples of human proteins exist where proteins exhibiting this motif demonstrate a non-ER distribution. Ones such example is the cytosolic and nuclear localisation of the KDEL-containing protein calreticulin (191). It remains unanswered how these proteins can evade ER retention, whether it is through an error rate of the KDEL receptor molecules, active secretion or whether KDEL-expressing proteins are released by necrotic cells. The implications of divergence from the canonical sequence results in altered affinity, in this case reduced, for the three known KDEL receptors which recognise the motif (192;193). As a result, the KTEL motif provides some plasticity in the subcellular localisation, and release from the ER, of AGR2 protein, which may be essential for dynamic protein function in response to physiological conditions (187;188). Recently, Gupta *et al.* have shown that the KTEL is essential for AGR2 for the AREG induction function of AGR2 (194), since KTEL deletion causes secretion and loss of function yet mutation of KTEL to KDEL does not rescue the function (188). This suggests that the KTEL motif is essential for AGR2 function, not only for conferring subcellular distribution.

Coupled to the N-terminal leader sequence and the KTEL motif, AGR2 demonstrates a single variant of the thioredoxin fold (TX), E/D-CXXS-Q motif which taxonomically describes the protein as being a member of the protein disulphide isomerase (PDI) family (Figure 1-5B). The presence of one to four classical TX-motifs facilitates oxidative/reductive reactions central to the appropriate shuffling of disulphide bonds during protein maturation (195). Curiously, the CXXS motif is divergent from the di-cysteine thioredoxin-fold, W-CXXC-K, exhibited by the founder ERP18 protein, and has been implicated in disease development through the covalent trapping of disulphide shuffling intermediates (196). This motif is not

restricted to AGR2 however, it is shared with two other PDIs/ER-proteins Eug1p and ERP44 (197;198). More recent research has described and characterised the existence of a substrate binding loop (amino acid residues 104-111) which forms the octa-residue docking site for the interaction between AGR2 and the AAA+ protein Reptin (199). Mutations in AGR2 at the junction codons of the loop at Phe-104 and Tyr-111 attenuate Reptin binding to the protein. This region maps to a comparable region of ERP18, and has been proposed to represent a substrate binding site for client proteins for the molecular chaperone function of ERP18 (196;200). During the scope of this thesis, truncated AGR2 (residues 41-175) has been structurally characterised by NMR (PDB code: 2LNS) and described as a homodimeric molecule, with a Kd of 8.83 μM, mediated through intermolecular salt bridges involving Gln-60 and Lys-64 (201). However, a further intermolecular disulphide interaction between the two cysteine-81 residues is reported (202;203), and thought to be necessary for the biochemical interaction with BiP/GRP78 (202).

Indeed, possible roles of AGR2 are complex since AGR2 subcellular localisation is highly disputed, with ER (179), nuclear (187;188), cell surface and extracellular distribution (204) all being reported. Human AGR2 is strongly expressed in tissues of the lung, stomach, colon, prostate and small intestine, organs which contain mucus secreting cells, or function as endocrine organs (Table 1-1) (205). AGR2 may also be involved in epithelial barrier function as the AGR2 promoter regulators are grouped into a family which are typical for epithelial goblet cells (171). AGR2 is also published to be expressed in the fetal liver, and the expression pattern conserved in the adult (206); with the protein present in tall columnar but not cuboidal cells of the intrahepatic, hilar and extrahepatic biliary tree (207). The differentiation of biliary epithelial cells lining bile ducts requires the acquisition of the columnar and mucus secretory phenotype (207). These observations have let to the assumption that AGR2 may have a role in the differentiation of biliary tree cells.

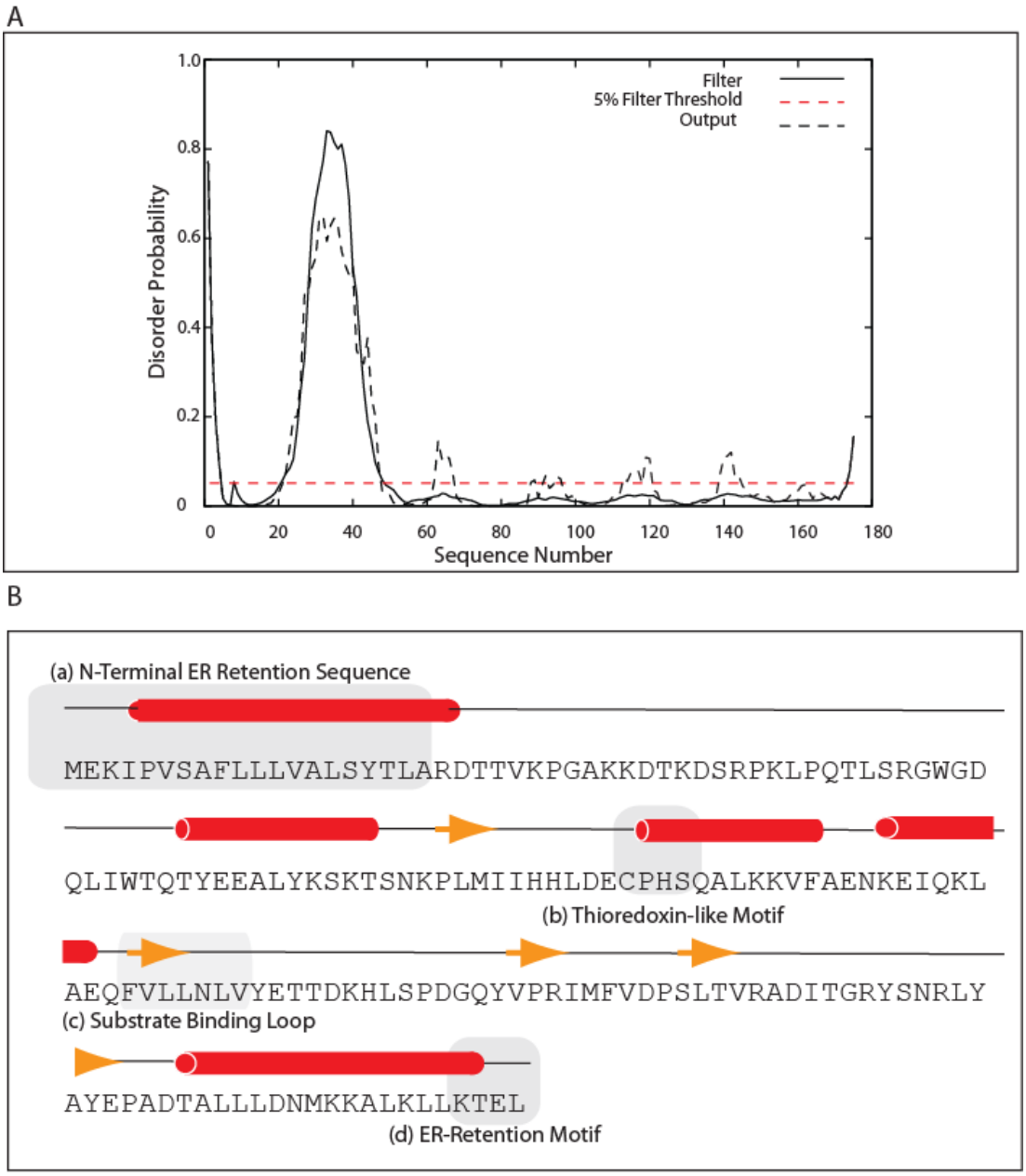


Figure 1-5 Structural determinants of human AGR2. (A) The N-terminal 45 amino acid residues indicate a high level of predicted protein disorder. (B) Schematic illustration of motifs present in the structure of AGR2, highlighting (a) a putative cleavable hydrophobic N-terminal signalling sequence, (b) the diverged thioredoxin-like CXXS motif, (c) the substrate binding loop fine mapped for the interaction of AGR2 with Reptin (199) (d) ER-retention motif recognised by KDEL receptors in the Golgi apparatus for retrograde transport to the Endoplasmic reticulum.

Table 1-1 The expression of AGR2 in normal human tissues. Immunohistochemical staining is grouped into high (red), moderate (grey) and weak (green) categories. Data derived from the Human Protein Atlas (http://www.proteinatlas.org/) and Brychtova *et al*. (153)

Normal Tissue Staining by Immunohistochemistry	Cell Type	Protein Expression
Appendix	Glandular cells	Strong
Breast	Glandular cells	Weak
Bronchus	Respiratory epithelial cells	Strong
Cerebral cortex	Neuronal cells	Moderate
Cervix, uterine	Glandular cells	Strong
Colon	Glandular cells	Strong
Corpus, uterine	Glandular cells	Strong
Duodenum	Glandular cells	Strong
Epididymis	Glandular cells	Strong
Oviduct	Glandular cells	Strong
Gall Bladder	Glandular cells	Strong
Kidney	Cells in tubules	Moderate
Lung	Alveolar cells	Moderate
Nasopharynx	Respiratory epithelial cells	Strong
Pancreas	Exocrine glandular cells	Moderate
Placenta	Trophoblastic cells	Moderate
Prostate	Glandular cells	Strong
Rectum	Glandular cells	Strong
Seminal vesicle	Glandular cells	Strong
Small intestine	Glandular cells	Strong
Stomach	Glandular cells	Strong
Tonsil	Squamous epithelial cells	Moderate
Urinary Bladder	Urothelial cells	Strong

1.3.4 Core biochemical functions of AGR2

Research into the structural elements and subcellular distribution of AGR2 has highlighted possible functions for evaluation. Further, several groups have, using a variety of techniques, begun to map the interactome of AGR2 protein (Table 1-2). These data are presented and discussed, supported by further studies, to outline the current understood biochemical roles of AGR2 (Figure 1-6).

Table 1-2 The interactome of AGR2. Data collated from literature indicating all of the published protein-protein interactions of AGR2. Techniques are abbreviated to Y2H (yeast-2-hybrid) and CoIP (co-immunoprecipitation).

Gene name	*Function*	*Method*	*Reference*
DAG1	Links the cytoskeleton and the extracellular matrix	Y2H	Fletcher *et al.* (2003) (172)
C4.4	Metastasis-associated GPI-anchored protein	Y2H	Fletcher *et al.* (2003)(172)
PROD1	Axotolt homolog for human CD59	Y2H	Kumar *et al.* (2007) (154)
KDELR	KDEL receptors	CoIP	Raykhel *et al.* (2007) (192)
MUC2	Mucin 2	CoIP	Park *et al.* (2009) (165)
RUVBL2	AAA^+ ATPase—DNA repair and transcription	Y2H	Maslon *et al.* (2010) (199)
BiP/GRP78	Unfolded Protein Response	CoIP	Ryu *et al.* (2013) (202)
HECTD1	E3 ubiquitin ligase	CoIP Y2H	Yu *et al*., (2012) (208) Unpublished data (169)
ARHGAP29	GTPase activating protein (Rho)	Y2H	Unpublished data (169)
CKAP2	Cytoskeletal-linked protein involved in mitosis	Y2H	Unpublished data (169)
CHD6	Chromatin remodelling factor	Y2H	Unpublished data (169)
GPSM2	Regulates G-protein activation	Y2H	Unpublished data (169)
HIVEP1	DNA binding protein	Y2H	Unpublished data (169)
NRIP1	Binds hormone-dependent receptors	Y2H	Unpublished data (169)
NRXN3	Controls adhesion and receptor signalling	Y2H	Unpublished data (169)
LYPD3	Regulates cell migration	Y2H	Unpublished data (169)

1.3.4.1 Endoplasmic reticulum related functions of AGR2

The most conspicuous motif of AGR2 that has been used in the taxonomic classification and therefore subsequent studies of function is the presence of the thioredoxin-like fold identified by Persson *et al.* (209). Coupled to the degenerative ER retrieval motif, ER localisation has been speculated and reported (179). The endoplasmic reticulum contains a network of chaperones facilitating the appropriate maturation and folding of proteins targeted for the secretory pathway (210). The oxidoreductases, protein disulphide isomerases (PDIs) and ER proteins (ERPs) are key components of this process (195;211). The PDI class of proteins are essential in many biological processes. These proteins are microenvironmentally regulated to act as enzymes to catalyse disulphide bond formation and isomerisation in newly synthesised cysteine containing substrates, preventing the formation of illicit intermediates and subsequent aggregated proteins. The presence of the canonical active motif residues, CXXC, contains two free thiol groups at each of the cysteine residues allowing reduction, shuffling and oxidation of disulphide bonds (212). PDI/ERP family members contain one to four of these TX domains (195;213;214) suggesting the complexity, and specificity of the ER folding machinery (209). The CXXS motif of AGR2 possesses lower activity associated with disulphide bond reorganisation, but may contribute to the isomerisation of existing disulphide bridges allowing the completion of specialised function (215;216). If PDI substrates are not appropriately folded, mis-folded proteins may accumulate in the ER instigating ER stress and eventually cell apoptosis (217). Although thioredoxin activity of AGR2 has never been observed *in vivo* or *in vitro* (169), AGR2 has been implicated in the formation of mixed disulphides in the maturation of cysteine rich glycoprotein, intestinal mucin, MUC2 (165) and MUC1 (180).

Recently, AGR2 has been described as performing a pivotal role in the control of endoplasmic reticulum homeostasis (179). In this study, AGR2 was identified bound to newly synthesised cargo proteins in a proteomic analysis of isolated ER-bound ribosomes. This indicated that AGR2 was associated with nascent protein chains co-translationally, when they were translocated to the ER lumen. Further, it was shown that AGR2 expression could be controlled by the unfolded protein response (UPR),

and that silencing of AGR2 disrupts the components of the ER-associated degradation machinery, resulting in a reduction in cells ability to cope with acute ER stress (179). Complementary to this, data suggest AGR2 induction in pancreatic tumour cells regulate the expression of several other ER chaperones, namely PDI, CALU and RCN1, proteins of the ubiquitin-proteosome degradation pathway (HIP2, PSMB2, PSMA3, PSMC3 and PSMB4) and lysosomal proteases (cathepsin B and cathpsin D) (204). Further mechanisms incorporating AGR2 to the UPR signalling pathway arises in the recent publication by Ryu *et al.* who describe AGR2 forming a homodimer, via an intermolecular disulphide bond, and show it to be essential in the interaction with BiP/GRP78, and resultant attenuation of ER stress induced cell death (202).

1.3.4.2 AGR2 as a cell growth effector

AGR2 over expression has been described in multiple human cancers (discussed in detail in 1.3.6.1). *In vitro* and *in vivo* studies have demonstrated that AGR2 promotes tumour growth, and AGR2 expression induces anchorage independent growth in soft agar assays (164;218;219), yet the mechanisms whereby AGR2 exhibits these affects is not fully appreciated.

One hypothesis is that AGR2 can induce the expression of the epidermal growth factor receptor (EGFR) ligand amphiregulin (AREG) via the Hippo pathway co-activator Yap-1 (194). AREG has previously been implicated as a growth promoter through *in vitro* and *in vivo* studies (220;221), and can interact with EGFR, albeit with lower affinity than epidermal growth factor (EGF) itself (222), activating the EGF pathway to stimulate cell growth, proliferation, and differentiation (223). In this study, when AGR2 levels were experimentally suppressed, exogenous AREG expression is able to rescue the transformed phenotype. Also, AGR2 does not function to activate any other known EGFR ligands and that activation resulted in increased phosphorylation of both EGFR itself and AKT downstream of the receptor. The Hippo pathway serves to regulate cell proliferation and apoptosis, and function in regulating organ size (224). Repression of the Hippo pathway leads to Yap-1

protein dephosphorylation and resultant translocation to the nucleus where it acts to inhibit apoptosis and promote cell division (225;226). Dong *et al.* (194) propose that AGR2 expression acts to dephosphorylate Yap-1, and subsequent induction of AREG, however, the specific molecular mechanism by which AGR2 achieves this is still under study. The coupling of AGR2 to both the Hippo and EGF signalling pathway however, gives an insight into the complexity of AGR2 function.

The second hypothesis is that AGR2 functions as an effector molecule to silence p53 growth suppressive function (150). The exogenous introduction of over expressed AGR2 into a cancer cell model can, while not altering cell-cycle parameters (150), (i) enhance cell survival in a clonogenic assay, similar to the loss-of-function p53 HIS175 allele (150), (ii) reduce p53 dependent transcriptional activity, using p300 activation as a readout, to a similar extent as the p53 regulator MDM2 (150), (iii) attenuate phosphorylation at p53 –activating phosphorylation sites (Ser15 and Ser392) following ultraviolet induced DNA damage (150), (iv) redistribute p53 such that 'active' p53 is excluded from the nuclear compartment, and localised to the cytosol in an 'inactive' environment (187)). In criticism, these studies were carried out in cell lines which do not normally express neither AGR2 nor p53, however a further study uncovered that 2,3,7,8-Tetrachlorodibenzo-p-dioxin (TCDD), a toxin and human carcinogen (227) that attenuates the p53 response to DNA damaging agents (228;229) induces AGR2 mRNA (six-fold) and protein (four-fold), further implicating the role of AGR2 in p53 inhibition (230). These data suggest that AGR2 functions as a survival factor promoting cell proliferation, and sequestering cells from p53-induced apoptotic events.

Transient silencing of AGR2 inhibits cell growth, cell-cycle progression and induces cell death (231). One study investigated the downstream intracellular signalling pathways following AGR2 suppression in a variety of breast cancer cell lines and identified modulation of cyclin D1, c-myc, and E2F1, supporting inhibitory effects on cell growth and cell cycle (231).

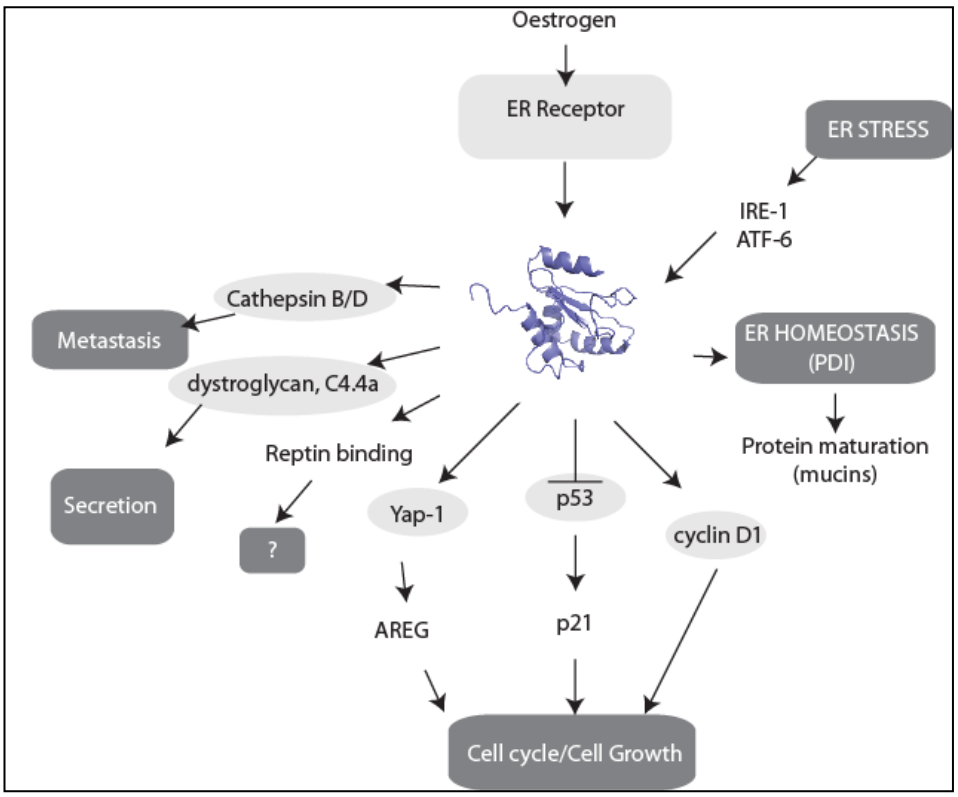


Figure 1-6 Schematic of the described regulation and functions of AGR2

1.3.5 Disease associations of AGR2

1.3.5.1 Implications of AGR2 overexpression in the aetiology of asthma

Aside from cancer, AGR2 gene has been shown to be over expressed following allergen exposure in murine models of asthma (232). Using qPCR and immunohistochemical validation it was determined that AGR2 is constitutively over expressed in the model, and primarily localised to the goblet cells of the lung. Subsequently, based on the accepted role of AGR2 in the maturation of intestinal MUC2 (165), AGR2 was analysed to investigate whether there was a relationship in the synthesis of the airway mucins MUC5AC and MUC5B (233). AGR2 was shown to be co-expressed in the same cells as MUC5AC and MUC5B, and was shown to co-immunoprecipitate with MUC5AC. *agr2 -/-* mice were used to determine the role of AGR2 in allergic airway disease, where it was elucidated that allergen challenged AGR2-null mice had >50% suppressed expression of the respiratory mucins compared to endogenously expressing wild-type mice (233). The reduction of secreted mucin was accompanied by increased proportion of mucin contained within the ER of secretory cells and evidence of ER stress in the airway epithelium. It is therefore assumed that AGR2 overexpression in asthma allows maturation and overproduction of respiratory mucins MUC5AC and MUC5B. As a result, mucus hypersecretion causes airway occlusion which is a major pathophysiologic feature of asthma (234;235). In a further study in human lung epithelial cells, the authors discovered that interleukin-13 (IL-13) silencing had a downstream effect on AGR2 and MUC5AC levels (236), and a murine model study agreed that AGR2 protein expression is significantly higher in the asthmatic model. Also, MUC5AC and AGR2 were discovered to be induced by IL-13, drug induced silencing of IL-13 led to a decrease in the expression of both AGR2 and MUC5AC, suggesting that IL-13 may perform an upstream regulatory role in this particular pathway (237) and implicates AGR2 in a new model of inflammatory respiratory disease.

1.3.5.2 Association of AGR2 with inflammatory bowel disease

The chromosomal location of AGR2, 7p21.3 (170), maps to a key susceptibility region of the human genome for inflammatory bowel disease (238;239). The

described role of AGR2 in mucus production led to its evaluation as a candidate gene for inflammatory bowel disease (171). It was discovered that single nucleotide polymorphisms (SNPs) in the 5' region of AGR2 are associated with an increased susceptibility for ulcerative colitis (171). As highlighted previously, studies have subsequently found that AGR2 expression is essential in goblet cell function and development (166), and that AGR2 knockout in murine studies disrupted the production of protective mucin (165). These mice suffer the onset of diarrhoea and increased risk of ulcerative colitis (165). The intestines of patients suffering from inflammatory bowel disease often exhibit significant symptoms of ER stress (240). The role of AGR2 in the secretory pathway, and relief or ER stress, reviewed above, support the finding that in normal functioning of the cell controlled AGR2 expression is a requirement for the appropriate maturation, localisation and secretion of extracellular glycoproteins. Dysfunction, through mutation, over expression and under expression has indicated significant pathological effects.

1.3.6 Cancer and cancer related pathways

1.3.6.1 AGR2 overexpression in human cancer

AGR2, and its close homologue AGR3, protein have been documented as being over expressed in a wide range of human cancers involved in both the onset and in progression of neoplasia (Table 1-3). Primarily concerning hormone dependent cancers such as breast (172;175;176;178;184;205;219;241-244), prostate (177;205;245;246) and ovarian (247-250), but also significantly in non-hormone dependent cancers such as tumours of the oesophagus (150;251), gastro-intestinal tract (252) and lungs (27;253). As described previously, AGR2 exhibits the basic features of an onco-protein, due to pro-growth and cell survival enhancing activity (150;175). Intriguingly, deletion of the 10 C-terminal amino acid residues, those harbouring the ER-retention site, prevented clonogenic growth stimulation (175) indicating a role in tumourigenesis of the ER localised protein and survival signalling. Coupled to this, the inhibition of p53 (150), and the promotion of p53 nuclear exclusion (187) in UV-induced DNA damaged cells, presents a candidate protein whose role in tumourigenesis warrants in depth study.

The role of AGR2 in cancer was first correlated to survival in a proteomics screen of pre-malignant lesions from Barrett's epithelium, a pre-neoplastic oesophageal disease caused by damage from acid reflux providing a selection pressure for survival factors leading to the transformation of normal cells (150;254). Polypeptides from AGR2 were found to be significantly overexpressed compared to normal oesophageal tissue, suggesting some role of AGR2 in the stress response of the cell, and further validated as influencing p53 activity. AGR2 had previously been identified as co-expressed with the oestrogen receptor in breast cancer cell lines (176), and it was shown that the *agr2* gene was responsive to oestrogen signalling (175).

In breast cancers, AGR2 is associated with poor prognosis (175;241) and a malignant metastatic phenotype (241). In ovarian cancers, mucinous cancer biopsies show particular increased AGR2 expression by immunohistochemistry (248;249), but also in endometriod cancers (247;255), and predicts poor prognosis in high-grade serous tumours (256). A high concentration of AGR2 in the extracellular plasma has been associated with both serous and non-serous ovarian cancer (247). High expression levels of AGR2 are also reported in prostate carcinomas (257;258), both intracellularly and secreted into the blood (259) and urinary secretions (245) of patients, and studies are now ongoing as to whether AGR2 could be used as a minimally invasive prognostic marker of prostate disease. Recent reports suggest that the FOXA1 and FOXA2 transcription factors are likely to be involved in AGR2 expression in prostate cancer, coupled to the inverse correlation of AGR2 levels with ErbB3 binding protein 1 (EBP1), an endogenous negative regulator of androgen receptor signalling, which may allow the invasive character of prostate tumours (260;261). AGR2 is associated with poor survival of prostate cancer patients (246), and silencing of AGR2 in prostate cancer cell lines redact the survival enhancement, allowing cells to enter senescence (262).

Despite the androgen-receptor activation of AGR2 levels, elevated levels of AGR2 have also been linked to non-hormone related oncogenesis and cancer progression. Levels of AGR2 mRNA have been found to be significantly enhanced in pancreatic cancer tissue compared to paired normal tissue; a 14-fold increase is reported (218).

This study is complemented by further investigation describing AGR2 as being a cofactor in the promotion of initiation and progression of pancreatic intraepithelial neoplasia (180). AGR2 protein was also detected in the pancreatic juices of pre-malignant neoplasia of the pancreas (263;264). AGR2 elevation is also reported in gastric cancer cells (252) and in colorectal cancer (265). In lung adenocarcinoma, AGR2 was isolated from the serum of patients (266), to such an extent that 94% of lung adenocarcinomas exhibited upregulated AGR2 expression (267). In non-small cell lung adenocarcinomas this trend continues (253;268).

Table 1-3 Correlation of AGR2 expression in human cancer

Cancer Type	AGR2 Manifestation	Reference
Breast (adenocarcinoma)	Overexpression/Upregulation	Fletcher *et al*., 2003; Duran *et al*., 2008
	Associated with ERBB2 tumours	Duran *et al*., 2008
	Poor Prognosis	Barraclough *et al*., 2009; Hrstka *et al*., 2010
	Metastasis	Fletcher *et al*., 2003; Liu *et al*., 2005; Barraclough *et al*., 2009
Colorectal	Poor prognosis	Valladares-Ayerbes *et al*., 2012
Gastric	Metastasis	Lee *et al*., 2011
Lung	Overexpression	Fritzsche *et al*., 2007; Chung *et al*., 2012
Oesophageal (Barrett's Premalignant)	Overexpression	Pohler *et al*., 2004
Ovarian (mucinous)	Upregulation	Park *et al*., 2011
	Overexpression	Gray *et al*., 2012
	High plasma concentration	Edgell *et al*., 2010
Ovarian (other)	Poor prognosis	Darb-Esfahani *et al*., 2012
	Overexpression/Upregulation	Armes *et al*., 2013
Pancreatic	Overexpression/Upregulation	Barry *et al*., 2012; Ramachandran., 2008)
	Secreted into pre-malignant neoplastic juice	Chen *et al*., 2010; Makawita *et al*., 2011
	Promotes initiation/progression of pancreatic epithelial neoplasia	Norris *et al*., 2012
Prostate	Overexpression in serum	Kani ei al., 2013
	Overexpression/Upregulation	Kristiansen *et al*., 2005; Bu *et al*., 2011
	Metastasis	Hu *et al*., 2012
	Poor prognosis	Zhang *et al*., 2007

1.3.6.2 AGR2 in metastasis

A key nature of AGR2 highlighted in many studies is, not only the effect of AGR2 in growth promotion (Section 1.3.4.2), but also the detection of AGR2 in excretory systems such as blood serum and urine, suggesting that AGR2 may act in a paracrine or autocrine manner. This prometastatic role of AGR2 is broadly accepted. Since its detection in a gene expression screen of cancer cell lines identifying candidate markers for metastasis (269), subsequent interactomic studies have demonstrated that AGR2 directly interacts with two transmembrane proteins involved in cell-cell contact or cell-matrix interactions in C4.4A and alpha-dystroglycan (Table 1-2) (172). Transfection of AGR2 into a benign rodent mammary cell line transforms the cells to the metastatic phenotype, and it was suggested that AGR2 positive cells expressed enhanced adhesive properties (219). Subsequently, AGR2 knock down in both non small cell lung carcinoma and pancreatic cancer cells compromised anchorage independent growth *in vitro* and the growth of xenografted tumours *in vivo* (164;166). AGR2 expression has also been reported in the proliferation, migration and invasion of head and neck squamous cell carcinoma (270). Additionally, ovarian cancer indicates increased plasma concentrations of AGR2 in plasma (247), cell models of which overexpressing AGR2, demonstrate upregulated gene expression of proteins involved in cell proliferation, invasion and angiogenesis, and suppressed negative regulators of these processes (249). Further, AGR2 was also shown to have differential expression in metastatic gastric cancer cells, compared to non-metastatic cells (252). Finally, in stress conditions, resembling the tumour microenvironment, for example serum depletion and hypoxia, AGR2 was identified as upregulated in a screen together with other pro-survival, pro-invasive and angiogenic genes (178). The role of AGR2 in metastasis is of significant interest, as successful metastases requires activation of genes that promote cell survival in environments of pathophysiological stress, therefore the necessity to characterise this metastatic marker has emerging clinical significance.

1.3.6.3 Correlation of AGR2 expression and drug resistance

The clinical significance of AGR2 is heightened, by the evidence that AGR2 can mediate cytotoxic drug resistance in tamoxifen, doxorubicin and cisplatin-resistant tumours (Table 1-4). The use of the anti-oestrogen, tamoxifen, to treat human breast cancers is widespread; however intrinsic resistance to the drug is a major problem. Identification of oestrogen responsive genes which are induced by tamoxifen yet play a survival role is a key aim in the study of tamoxifen drug resistance (271). AGR2 expression is elevated rather than inhibited in response to tamoxifen treatment (175). In contrast, in cancer patients treated with the anti-hormonal, oestrogen suppressor drug Letrazole, the AGR2 gene was one of the top down regulated genes in biopsies from post-treatment patients who had responded well to the drug, suggesting that resistance to anti-oestrogens might correlate to failure to suppress AGR2 (175). Secondly, a label-free proteomics approach, PaCIFIC, was employed to screen for dominant proteins induced by tamoxifen treatment. AGR2 was highlighted as the protein whose expression was induced to the second greatest degree (174). At the molecular level, it was demonstrated that the tamoxifen induced expression of AGR2 was a direct effect of oestrogen receptor-α activation by chromatin immunoprecipitation. Elevated oestrogen receptor-α was bound to the AGR2 promoter in the presence of oestrodiol or tamoxifen and the AGR2 promoter was activated using a luciferase reporter assay (175). Recent data, using an RNA interference library, suggests that AGR2 is activated by the phosphoinositidine-dependent protein kinase 1-AKT (PDPK1-AKT) signalling pathway in response to tamoxifen treatment (272).

Proteomic analysis studies, supported by cell based MTT assay, have also implicated AGR2 over expression in resistance to the DNA damaging agents doxorubicin and cisplatin *in vitro* (174;175;248). AGR2 has also been implicated in anti-oestrogen resistance comparing fulvestrant-sensitive and –resistant T47D cell lines (273) Conversely, AGR2 expression is reduced in docetaxel-resistant tumours (274). Ongoing studies have begun to develop tools for anti-AGR2 therapeutic leads and detection of AGR2 in clinical tissue samples for diagnostics, including the publication of an AGR2 interacting peptide aptamer (TXIYY) (187;275).

Table 1-4 Studies highlighting AGR2 in therapeutic cancer drug resistance

Drug	Role of AGR2 in Resistance	Reference
Tamoxifen	Induced by tamoxifen, Overexpression in resistant tumours	Hrstka *et al*., 2010; Hengel *et al*. 2011
Cisplatin	Mediates resistance in xenograft models	Hengel *et al*. 2011; Gray *et al*., 2012
Doxorubicin	Overexpressed in resistant tumours	Hrstka *et al*., 2010; Hengel *et al*. 2011
Docetaxel	Reduced expression in resistant tumours	Zhao *et al*., 2009
Fulvestrant (Anti-oestrogen therapy)	Upregulated in resistance cell line (with cathepsin D)	Huber *et al*., 2004

1.3.7 Anterior Gradient-3

The chromosomal location of Anterior Gradient-3, also known as AGR3, HAG-3 and BCMP-11, is contiguous with AGR2 at 7p21.1-3. AGR3 shares 71% sequence similarity with AGR2 (Figure 1-7) and the expressed product shares the CXXS core motif thought to be required for disulphide bond formation (209). AGR3 was first identified as a protein present in breast cancer membrane fractions (184). Follow-up studies indeed identified AGR3 as co-expressed with AGR2 in breast cancer tissues with a strong correlation with oestrogen receptor α status (172). Yet expression of AGR3 is not coupled to AGR2 in prostate cancers indicating a lack of redundancy. It is necessary to indicate, that gene/protein studies of AGR2 highlighted, AGR3 was not found to be induced by oestrogen (176) (however in recent studies androgen-responsiveness has been reported in prostate cancer cells (276)) or tamoxifen (174), or as an inhibitor of p53 (150;275), thus there is no indication whether AGR3 shares the oncogenic nature attributed to AGR2. As a result, there has been limited data published on AGR3, despite the interest in AGR2 in health and disease. We have previously published data investigating the expression of AGR3 in ovarian cancer (248), to identify a suitable clinical model for the study of AGR3 function. Through

the development of well characterised monoclonal antibodies specific for the AGR3 protein, and not AGR2, we identified that (i) AGR3 does not co-localise to the endoplasmic reticulum to the same extent as AGR2 protein, (ii) AGR2 is over-expressed in four different types of human primary ovarian cancer, (iii) AGR3 and AGR2 exhibit coupled expression in mucinous ovarian cancer, but uncoupled in 3 other types of primary ovarian cancer and (iv) that AGR3 can mediate drug resistant growth in an isogenic cancer cell panel (248).

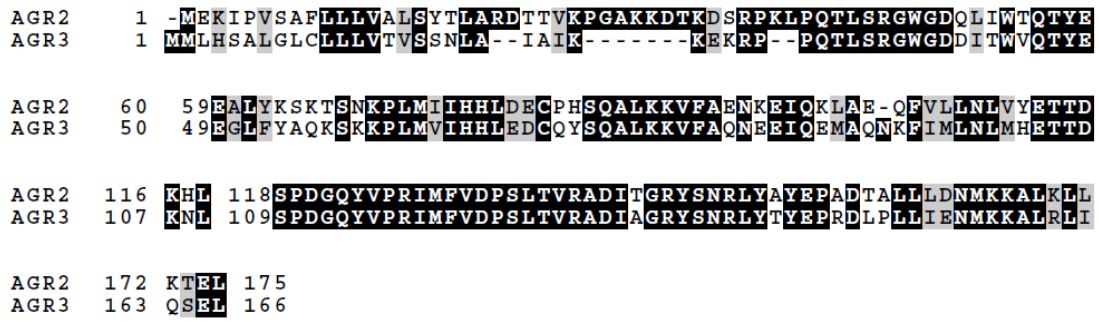


Figure 1-7 Sequence alignment of human AGR2 and AGR3 indicating high levels of similarity and conservation of protein sequence. Alignment carried out using ClustalW2 (http://www.ebi.ac.uk/Tools/msa/clustalw2/) and shading by BoxShade (http://www.ch.embnet.org/software/BOX_form.html).

1.4 Aims

The aims of these studies were to broadly investigate the hypothesis of Anterior Gradient proteins as novel oncoproteins. This was targeted through the engineering of a well characterised isogenic cell line panel providing reagents for the study of the effect of the expression of the Anterior Gradient genes in cell growth and the response to drug induced DNA damage (Chapter 3). Subsequently, these experimental cells could be examined for consequential changes at the transcriptional level through gene expression studies, and thorough analysis of changes in the dynamic protein landscape. Proteomic studies, utilising novel mass spectrometry methodologies coupled with bioinformatical pathway analysis tools, were applied to gain information regarding the previously unclear biochemical functions of AGR2, in terms of normal cellular function and the connection to oncogenesis, prior to comprehensive biochemical validation of identified cofactors (Chapter 4).

Structural approaches were also undertaken to identify the functional role and bioactive modulators of the recently discovered homodimeric complex of AGR2 (201-203) (Chapter 5). This required the development and optimisation of reagents for, and the inception of, a novel two site microplate format assay targeted at describing the oligomeric status of AGR2. Exploitation of this assay allowed the confirmation of structural motifs of the protein which can allosterically destabilise/stabilise the complex and others which are essential for the interacting interface of the dimer. By means of a library of natural compounds, potential leads for compounds disrupting of the quaternary protein structure were highlighted.

Chapter 2: Materials and Methods

2.1 Reagents and chemicals

All reagents were purchased from Sigma Aldrich unless otherwise stated. Tissue culture reagents including Dulbecco's modified eagle medium (DMEM), Roswell Park Memorial Institute (RPMI) and Trypsin were purchased from Gibco, Life Technologies. Foetal bovine serum (FBS) was sourced from Biosera.

2.2 Equipment

Sorvall RC-5C plus and Eppendorf 5415R were used for all centrifugations. DNA and RNA concentrations were measured using a NanoDrop spectrophotometer (Thermo Scientific) at A_{260}. Bradford protein determination assay was read using a PerkinElmer Victor^3 1420 multi-label plate reader with 595 nm filter. SDS-PAGE was carried out using the BioRad Protean II mini-gel system. Gels containing ^{35}S radioactive isotope labels were visualised through the use of Phosphoimager (Storm 840, Amersham Biosciences). Enhanced Chemiluminescence (ECL) based ELISA experiments were read using a Fluroskan (Ascent FL). 800 nm fluorophore-conjugated antibodies were detected using the 800 nm channel of an Odyssey SA plate reader (Licor). PCR was performed using a DNA Engine Dyad Peltier thermal cycler (BioRad) and quantitative real-time PCR using MJ Research PTC-200 Peltier thermal cycler (BioRad).

2.3 Microbiological Procedures

2.3.1 Growth of bacterial cultures

Overnight cultures were prepared with 5-200 mL of LB broth media supplemented with the appropriate selective antibiotic (100 µg/mL ampicillin or 50 µg/mL kanamycin) and inoculated with colonies from either a glycerol stock or single colony isolation from a bacterial culture dish. The cultures were incubated overnight at 37°C with constant agitation (220 rpm). Sterile culture vessels were at least four times the volume of culture to allow sufficient aeration.

Luria-Bertani (LB) broth

1 % (w/v) Tryptone

1 % (w/v) NaCl

0.5 % (w/v) Yeast extract

Dissolved in distilled water and sterilised by autoclaving at 121°C for 20 minutes

To prepare the culture plates the LB agar plates, 1.5% (w/v) bacto-agar was added to LB media. LB agar was melted in a microwave oven and allowed to cool sufficiently. Cooled agar was supplemented with appropriate selective antibiotic (50 µg/mL kanamycin or 100 µg/mL ampicillin) and poured into 100 mm Petri dishes (Sterilin) with a convective air current from a Bunsen burner, and left to cool. Culture plates were stored at 4°C for no longer than one month. Subsequently plates were warmed to 37°C for 1 hour prior to plating.

2.3.2 Glycerol stocks

For long term storage of the bacterial cultures, glycerol stocks were prepared. 800 µL of an overnight bacterial culture was mixed thoroughly with 200 µL of sterile

80% glycerol, transferred to a cryotube (Nunc), snap frozen in liquid nitrogen and stored at -80°C.

2.3.3 Preparation of competent cells

Bacterial cells were inoculated into 5 mL of LB and incubated overnight at 37°C with shaking at 220 rpm. The culture was diluted 1:200 into 100 mL of LB and incubated at 37°C until the optical density at 600 nm ($OD_{600\ nm}$) reached 0.6. Cells were centrifuged (20 minutes at 4°C, 4000 rcf) to pellet prior to being resuspended in 15 mL of ice cold Buffer 1 and incubated on ice for one hour. Cells were centrifuged again, as above, and pellet resuspended in 4 mL of ice cold Buffer 2 before incubation on ice for 15 minutes. Cells were aliquotted, in 50 µL volumes, into pre-chilled microcentrifuge tubes, snap frozen in liquid nitrogen and stored at -80°C.

Buffer 1	*Buffer 2*
100 mM $RbCl_2$	10 mM MOPS pH 6.8
79 mM $MnCl_2$	10 mM RbCl
30 mM CH_3COOK pH 7.5	13.5 mM $CaCl_2$
13.5 mM $CaCl_2$	15% Glycerol
15% (v/v) Glycerol	pH adjusted to 6.8 and filter sterilised
pH adjusted to 5.8 and filter sterilised	

2.3.4 Heat shock transformation of *E. coli*

50 µL of competent cells (BL21 or DH5α *E. coli*, stored at -80°C) were thawed on ice. Bacteria were inoculated with 100-500 ng of plasmid DNA and incubated on ice for 30 minutes. The bacteria were heat-shocked at 42°C for 45 seconds in a water bath or heat block and returned to ice. Bacterial cell culture was then made up to 1

mL with LB media and cells recovered by incubation for 1 hour at 37°C (220 rpm). 200 µL of transformed cells were plated onto agar plates (supplemented with appropriate antibiotic) and incubated at 37°C overnight.

2.4 Molecular Biology Techniques

2.4.1 Amplification, purification and quantitation of plasmid DNA

An isolated bacterial colony was picked from a selective LB-agar plate and inoculated into 5 mL of LB, containing a selective antibody if required, and grown for several hours at 37°C with constant agitation at 220 rpm. Following incubation, the starter culture was diluted into 250 mL of LB (containing selective antibiotic if necessary) in a sterile flask and returned to the incubator for a further 16 hours. The bacterial culture was centrifuged at 4°C for 15 minutes at 6000 rcf to pellet and the supernatant discarded. Plasmid DNA was extracted from bacteria using plasmid DNA MaxiPrep kit (Qiagen) according to the manufacturer's instructions. If a lower yield of plasmid was required, plasmid DNA MiniPrep kit (Qiagen) could be used to extract DNA directly from the bacterial starter culture. Once purified, plasmid DNA was eluted from the purification column in nuclease-free water and stored at -20°C. The concentration of extracted DNA was quantified using the NanoDrop.

2.4.2 Agarose gel electrophoresis for the separation of DNA

Agarose gels were prepared by mixing electrophoresis grade agarose (Life Technologies) with 1× TAE buffer to a final concentration of 1% (w/v). The mixture was heated in a microwave oven to melt the agarose. The agarose was cooled, and inoculated with 1:10,000 dilution of SYBRSafe dye (Life Technologies). The solution was poured into a prepared agarose gel casting tank, comb inserted, and the gel allowed to set at room temperature. Once set, the gel was submerged in 1× TAE buffer. DNA samples were mixed with the 6× loading buffer, to a final

concentration of 1× of loading buffer. DNA standards were loaded (100 b or 1 kb ladder, NEB) and samples pipetted into wells of the gel, and the agarose gel was run at 100V for 60 minutes, or until bands were adequately separated. The bands were subsequently visualised under blue light (Safe Imager 2.0, Life Technologies) using Syngene (Genesnap).

1 × TAE Buffer	*6× DNA loading buffer*
40 mM Tris	0.25% bromophenol blue
1 mM EDTA	0.25% xylene cyanol FF
pH adjusted to 8.0	15% Ficoll

2.4.3 DNA Sequencing

DNA was sequenced using the Big Dye Terminator V3.1 Cycle Sequencing Kit (Applied Biosystems). The reaction was prepared as follows, then subject to thermal cycling as described:

Sequencing Reaction	*Thermal cycling conditions*
2 µL of 5× Sequencing Buffer Big Dye Terminator V3.1	1 - 96^{o}C for 1 minute
1 µL of Big Dye Terminator V3.1 Cycle Sequencing Kit	2 - 96^{o}C for 10 seconds
300 ng of DNA template	3 – 50^{o}C for 5 seconds
1 µL of sequencing primer (10 µM)	4 - 60^{o}C for 4 minutes
Final volume adjusted to 10 µL with nuclease-free water	5 – Repeat cycles 2-4 for 25 cycles
	6 – Hold at 4^{o}C

Sequencing primers used included stock primers of T7 forward, V5 reverse and M13 forward and reverse, as necessary from plasmid map. Following PCR, 2.5 µL of 125 mM EDTA and 30 µL of 100% ethanol were added to the sequencing reaction to precipitate DNA. The mix was vortexed and incubated for 15 minutes at room temperature. Samples were centrifuged for 20 minutes at 13200 rpm and the supernatant aspirated. The samples were briefly spun again to remove residual solvent. The DNA pellet was washed with 70% ethanol, and centrifuged for 5 minutes at 13200 rpm, and supernatant discarded before a final centrifugation and removal of remaining ethanol. The DNA pellet was then air dried and sequences analysed by Sequencing Service of Source Biosciences. Alternatively, at least 100 ng of purified plasmid DNA was sent in liquid form to the sequencing service that carried out an optimised protocol similar to above.

2.4.4 Cloning

2.4.4.1 Gateway cloning

One method of cloning the DNA sequence of the gene of interest into a plasmid was using the Gateway cloning system (Life Technologies). The manufacturer's protocol was followed. This required the PCR generation of the gene of interest flanked by attB sites, followed by a BP reaction into an entry clone (pDONR221), and finally the LR reaction into the expression vector. Briefly, primers (Sigma) were designed to incorporate the attB recombination sites into the PCR product (Table 2-1), and contained 18-25 bases corresponding to the 3' and 5' ends of the gene to be cloned, as below. Following PCR amplification and BP reaction the amplified gene was recombined into the pDONR221 backbone.

Table 2-1 Primers designed and synthesised for the generation of PCR fragments of genes on interest with attB1 and attB2 sites for Gateway cloning. Red: Poly-guanine required for attB recombination, Green: attB1 or attB2 sites, Orange: Kozak sequence (assists in the initiation of translation in eukaryotic cells), Black: gene sequence specific to gene of interest.

Primer Name	*Sequence (5'-3')*
AGR2 Gateway Forward	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGAAGGAGATAGAACCATGGAG AAAATTCCAGTGTC
AGR2 Gateway Reverse	GGGGACCACTTTGTACAAGAAAGCTGGGTCTTACAATTCAGTCTTCAGC
AGR3 Gateway Forward	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGAAGGAGATAGAACC ATGATGCTACACTCAGCTTTG
AGR3 Gateway Reverse	GGGGACCACTTTGTACAAGAAAGCTGGGTCTTATAGCTCTGACTGAATA

Using stock DNA constructs of AGR2-pDONR 221 and AGR3-pDONR 221 (a kind gift from Dr E. Murray), a Gateway LR reaction was prepared using 100 ng of entry clone, with 150 ng of the destination vector (pcDNA3.2/FRT/V5) and TE buffer adjusted to a final volume of 8 µL. 2 µL of LR clonase enzyme was added, mixed and incubated overnight at 25°C. Following the incubation, 1 µL of proteinase K solution was added and the mixture incubated at 37°C for 10 minutes to stop the reaction. 1 µL of the reaction was transformed into DH5α *E. coli* and plated on LB-agar plates supplemented with 100 µg/mL ampicillin. Single colonies could be isolated and grown in 5 mL LB supplemented with 100 µg/mL ampicillin prior to MiniPrep isolation of DNA. Plasmids were then sequenced using the T7 forward and the V5 reverse primers (Source Biosciences).

2.4.4.2 Conventional cloning using restriction enzymes

An alternate method of cloning of the sequence of interest into the required vector was using restriction enzymes and a ligation protocol. Briefly, the desired sequence was first amplified by PCR using primers flanked by appropriate restriction sites, followed by restriction enzyme digestion of both PCR product and vector and subsequent ligation of digested vector and insert.

Primers were designed to incorporate 18-25 bases specific to the N-terminal and C-terminal portion of the protein to be cloned, and the appropriate restriction sites chosen according to the multiple cloning site (MCS) present in the desired vector. It was sometimes necessary to incorporate additional nucleotide bases to ensure that the gene of interest was in frame relative to the plasmid backbone. Finally, primer sequences were modified to obtain a GC content of 40-60% and a melting temperature in the range of 60-80°C. A list of primers used is presented in Table 2-2.

Table 2-2 Primers used in cloning of $AGR2_{21-175}$ and AGR2-Δ45 into pEHISTEV highlighted in are the specific sequences recognised by restriction enzymes Nco1 (yellow) and Xho1 (red)

Vector	*Target*	*Primer sequence (5'-3')*
pEHISTEV	$AGR2_{21-175}$	Forward (Nco1):
		GGGCCATGGCTATGAGAGATACCACAGTCAAA
		Reverse (Xho1):
		CCGCCTCGAGTTACAATTCAGTCTTCAGCAACTT
pEHISTEV	AGR2-Δ45	Forward (Nco1):
		CCATGGCTATGAGAGGTTGGGGTGACCAAC
		Reverse (Xho1): CTCGAGTTACAATTCAGTCTTCAGCAACTT

PCR was used to amplify the gene from a template DNA source. The PCR reaction was set up as follows:

12 µL 2× Pfu Mastermix

2.5 µL Band Doctor (part of Pfu Mastermix kit)

50 ng Template DNA

1.25 µL Forward Primer (10 µM stock)

1.25 µL Reverse Primer (10 µM stock)

Adjusted to 25 µL with nuclease-free water.

The thermocycling conditions were prepared as follows:

1. 95°C for 2 minutes

2. 95°C for 20 seconds

3. 58°C for 40 seconds

4. 72°C for 2-5 minutes (1 minute per 2 kb)

5. Repeat steps 2-4 for 25-35 cycles

6. 72°C for 5 minutes

7. Hold at 4°C

Following PCR, 5 µL of the PCR product was separated on a 1% agarose gel to ensure amplification of a product of the correct size, and to confirm quality. The remaining PCR product was cleaned using PCR Clean-Up kit (Qiagen).

After clean up of the PCR product, both PCR product and vector were digested using restriction enzymes and buffers (NEB). Digest reactions were prepared as below:

Double digestion of vector	*Double digestion of PCR product*
1 µg of plasmid	30 µL of PCR product
5 µL NEB Buffer	5µL NEB Buffer
5 µL of BSA (if required)	5 µL BSA (if required)
1 µL Restriction enzyme 1	1 µL Restriction enzyme 1
1 µL Restriction enzyme 2	1 µL Restriction enzyme 2
Adjusted to 50 µL with nuclease-free water	Adjusted to 50 µL with nuclease-free water

To control for restriction enzyme inactivity, single digest reactions were also prepared to ensure that the enzymes were active under the conditions used. The digests were incubated for 3 hours at 37°C in a water bath or heat block. Following digestion, complete reaction volumes were supplemented with 6× loading buffer and separated on a 1% agarose gel. Bands were visualised under blue light and single bands corresponding to double digested vector and insert excised then extracted from the agarose gel matrix using Gel Extraction kit (Qiagen) according to manufacturer's instructions. DNA was eluted with nuclease-free water and quantified by NanoDrop.

Ligation reactions were then prepared to join together the digested insert and vector in the correct orientation. Using T4 ligase (Promega), and following the manufacturers guidelines, 100 ng of digested vector was use and the amount insert calculated using the following formula:

Insert (ng) =

vector (ng) × size of insert (kb) / (size of vector (kb) × molar ratio (insert/vector)

A 1:1 molar ratio of insert to vector was most commonly used.

The ligation reaction was prepared:

1 µL T4 Buffer

X ng insert

100 ng Vector

1 µL T4 ligase

Adjusted to 10 µL with nuclease-free water

For control purposes, conditions were set up which lacked insert to control for re-ligation of vector. The ligation reactions were incubated for 16-48 hours at 4°C. The complete 10 µL was transformed into *E. coli* DH5α and plated on LB-agar plates with selective antibiotic. Following selection, several colonies were picked, grown in 5 mL culture and subject to MiniPrep prior to plasmid sequencing.

2.4.4.3 Site-directed mutagenesis

Specific mutations were also cloned in plasmids to challenge the function of specific amino acid residues, or motifs, of the AGR2 protein. The desired change at the DNA level was made in synthesised primers (Sigma), before PCR amplification of the mutant gene within the plasmid. The primers used to make specific changes are presented below (Table 2-3):

Table 2-3 Primer sequences derived for site-directed mutagenesis codons highlighted in red are mutated from the template plasmid

Mutation	*Primer sequence (5'-3')*
FRT-STOP	Forward: ATGGAGAAAATT**TAG**GTGTCAGCA Reverse: TGCTGACAC**CTA**AATTTTCTCCAT
FRT-AGR2-KDEL	Forward: CTCAAGTTGCTGAAG**GAT**GAATTGTAGGACCC Reverse: GGGTCCTACAATTC**ATC**CTTCAGCAACTTGAG
FRT-AGR2-ΔKTEL	Forward: CTCAAGTTGCTG**TAG**ACTGAATTGTAGGACCC Reverse: GGGTCCTACAATTCAGT**CTA**CAGCAACTTGAG
AGR2-Y58A pEHISTEV	Forward: ATCTGGACTCAGACA**GCT**GAAGAAGCTCTATAT Reverse: ATATAGAGCTTCTTC**AGC**TGTCTGAGTCCAGAT
AGR2-E59A pEHISTEV	Forward: TGGACTCAGACATAT**GCA**GAAGCTCTATATAAA Reverse: TTTATATAGAGCTTC**TGC**ATATGTCTGAGTCCA
AGR2-E60A pEHISTEV	Forward: AGACATATGAA**GCA**GCTCTATATAAATCCAAG Reverse: CTTGGATTTATATAGAGC**TGC**TTCATATGTCT

A PCR reaction using the mutant primers was prepared as follows

12.5 µL 2× Pfu mastermix

2.5 µL 5× Band Doctor

50 ng template DNA

1.25 µL forward primer (10 µM)

1.25 µL reverse primer (10 µM)

Adjusted to 25 µL with nuclease-free water.

These reactions were subject to the following thermocycling program:

1. 95°C for 1 minute

2. 95°C for 50 seconds

3. 55°C for 1 minute

4. 68°C for 15 minutes

5. Repeat steps 2-4 for 19 cycles

6. 68°C for 30 minutes

7. Hold at 4°C

Following the completion of PCR amplification, 5 µL of reaction was removed for control purposes. 1 µL of DpnI restriction enzyme (NEB) was added to the PCR product reaction, and mixed. The digestion was incubated at 37°C for 2 hours to cleave methylated bacterial DNA (the template). After digestion the DpnI enzyme was heat-inactivated at 65°C for 10 minutes. 2µL of the digested product (and control undigested product) was transformed into *E. coli* DH5α cells and plated on LB-agar plates supplemented with appropriate selective antibiotic. Several colonies (3-5) were isolated, grown in 5 mL of LB and subject to MiniPrep DNA purification. Successful mutagenesis was confirmed by plasmid sequencing.

2.5 Tissue Culture Procedures

2.5.1 Cell lines and culture

Cell lines were incubated at 37°C and 5% CO_2 in a humidified incubator, with the exception of A375 cells which required a CO_2 saturation of 10% for optimal growth. Required media, RPMI or DMEM (Gibco, Life Technologies), was supplemented with 10% (v/v) FBS (Biosera) and 1% (v/v) penicillin/streptomycin mix (Life

Technologies). Table 2-4 lists the cell lines, their media requirement and AGR2 status, used in this study.

Table 2-4 Cell lines and tissue culture media requirements. Media used were Dulbecco's modified eagle medium (DMEM) and Roswell Park Memorial Institute (RPMI).

Cell Line	*Cancer Type*	*AGR2 Status*	*Media*
A375	Human skin melanoma	Negative	DMEM
MCF7	Human breast adenocarcinoma	Positive	DMEM
OE19	Human oesophageal adenocarcinoma	Positive	RPMI
A549	Human lung adenocarcinoma	Positive	DMEM
T47D	Human ductal breast epithelial	Positive	RPMI

2.5.2 Passaging of cells

Cell lines were grown in sterile 100 mm tissue culture plates (Corning) until near 100% confluence. To propagate growth, tissue culture medium was aspirated, and discarded. Cells attached to the substrate were washed twice with 10 mL sterile phosphate buffered saline pH 7.4 (PBS) prior to the addition of 3 mL trypsin-EDTA (Life Technologies) and incubation at 37°C in the incubator for approximately 5 minutes. Following incubation, cells could be observed detached from the culture vessel. 7 mL of fresh complete tissue culture media was added to the plate, and the suspension transferred to a 15 mL tube. Cells were appropriately diluted and seeded at the desired density into new sterile tissue culture plates as necessary.

2.5.3 Storage of mammalian cell lines and recovery

In order to maintain a low passage number of cells for reproducibility, cells could be prepared for long term liquid nitrogen storage. Cells were seeded into tissue culture plates and allowed to grow to 100% confluency. Following growth, media was discarded and the cells washed with PBS and trypsinised from the tissue culture substrate as previously. Once cells were detached, fresh media was added and the cells transferred to a 15 mL Falcon tube. The cells were centrifuged for 5 minutes at 1000 rpm, allowing the cells to pellet and the media discarded. Cells were resuspended in 5 mL of cell freezing solution (10% DMSO in FBS), and 1 mL aliquots prepared in cryovials. Vials were then transferred to an isopropanol bath (Nalgene) and incubated at -80°C for 24 hours. Once gradually frozen, the aliquots were moved to liquid nitrogen for permanent storage.

To recover cells, vials were rapidly thawed at 37°C in a water bath. Thawed cells were then plated into a tissue culture plate with 10 mL of fresh media and incubated in the tissue culture incubator. 24 hours later, the media was aspirated and exchanged to remove traces of DMSO. The cells could then be grown to confluency prior to experimental seeding.

2.5.4 Transfection of plasmid DNA

The cell line to be transfected was seeded in 100 mm tissue culture plates and allowed to grow to 75-90% confluence. Media was discarded and fresh added prior to transfection. The transfection mixture was prepared, as below, incubated for 15 minutes at room temperature then added to cells drop wise. The cells were incubated for 24-48 hours prior to harvesting. The reaction could be scaled for tissue culture vessels of smaller surface area. For control purposes, additional cells were transfected with the specified plasmid lacking the gene of interest. Plasmids used for transfection are highlighted in Table 2-5.

Plasmid Transfection Preparation

0-10 µg Plasmid DNA

15 µL Attractene (Qiagen)

Made up to a final volume of 300 µL with serum free media.

Table 2-5 List of plasmids used for cell transfection

Plasmid	*Source*
wtAGR2-pEF5/FRT/V5-DEST	Cloned in-house
AGR2-KDEL-pEF5/FRT/V5-DEST	Cloned in-house
AGR2-ΔC-pEF5/FRT/V5-DEST	Cloned in-house
TSG101-pCMV-XL5	Origene

2.5.5 Hygromycin B selection of stably transfected cells

In order to stably transfect A375 (AGR2-null) cells with the single gene insert of the AGR2, the Flp-In system (Life Technologies) was utilised. Briefly, cells were co-transfected with the required plasmid and pOG44 plasmid at a ratio of 1:9 (0.5 µg + 4.5 µg or 1 µg + 9 µg, respectively) and Attractene, as Section 2.5.4. As a control, one condition was prepared with pOG44 only. 24 hour post-transfection, media was removed and replaced with fresh media supplemented with 50 mg/mL hygromycin B (Sigma). Hygromycin supplemented media was exchanged every 2-3 days for approximately 2 weeks, until the control transfected cells (lacking the gene of interest and hygromycin resistance gene) had died and colonies were evident in the experimental plates. Colonies were trypsinised from the plate, transferred to new tissue culture vessels and grown to confluency.

2.5.6 Transient silencing of gene expression using siRNA

Cells were seeded in 6-well plates and grown to 40-60% confluence for siRNA-induced gene silencing. siRNA against AGR2 (ON-TARGET*plus*, Thermo Scientific) and TSG101 (siGENOME SMARTpool, Thermo Scientific) was used at a final concentration of 50 nM. For each well, 5 µL of 20 µM siRNA stock was diluted using serum free, and antibiotic free media, to a final volume of 200 µL, concurrently, 4 µL of DharmaFECT I reagent was diluted with 196 µL of serum and antibiotic free medium. Each mixture was prepared and incubated at room temperature for 5 minutes, prior to the diluted siRNA and DharmaFECT reagents being combined and incubated for a further 20 minutes. During the incubation, tissue culture media was aspirated from the cells in culture, and replaced with 1.6 mL of fresh media. Subsequently, the entire 400 µL siRNA-DharmaFECT reaction was added to the cells and incubated for 24-48 hours in the tissue culture incubator. Cells were then harvested or treated as per figure legends.

2.5.7 Drug treatment

Drug compounds were used to analyse the response of experimental cells. Cells were seeded in 6 well or 100 mm tissue culture plates prior to the dilution of drug to appropriate concentration (Table 2-6).

Table 2-6 Drug compounds used and details of treatment

Drug Compound	*Final Concentration*	*Function*	*Incubation*	*Supplier*
Cisplatin	20 µM	DNA damaging agent	1-24 hours	Sigma
Polyinosinic:polycytidylic acid (poly I:C)	25 µg/mL	TLR3 activation	1-6 hours	Source Bioscience
Tunicamycin	1 µg/mL	ER stress induction	1-4 hours	Sigma
MG132	10 µM	Inhibition of proteosomal recycling	3 hours	VWR

2.5.8 Harvesting cells

Tissue cultures were placed on ice in preparation for harvesting. Growth media was discarded prior to cells being washed twice with ice cold PBS. Cells were scraped into ice cold PBS (1 mL for 100 mm plate, 400 µL for 6 well plate) using a cell scraper, before transferring to a pre-chilled microcentrifuge tube. Cells were pelleted by centrifugation at 5000 rpm for 5 minutes at 4^0C. The supernatant was aspirated, and the cells snap frozen in liquid nitrogen. Cells could then be stored at -80°C for later analysis or resuspended directly in lysis buffer for analysis.

2.5.9 Cell lysis

Pelleted cells were thawed on ice, before resuspension in approximately 3× pellet volume of RIPA cell lysis buffer. Lysis was carried out for 20 minutes on ice with periodic vortexing. The lysate was then centrifuged at 13200 rpm for 12 minutes at 4°C to pellet insoluble material. The supernatant was collected, transferred to a fresh microfuge tube and protein concentration assessed by Bradford assay. The lysate

could be immediately used for downstream analysis or snap frozen in liquid nitrogen for -80°C storage.

RIPA lysis buffer

1% Triton-X100

1% Sodium deoxycholate

0.1% SDS

0.15 M NaCl

0.01M Sodium phosphate pH7.2

1× Protease inhibitor (Roche)

2.5.10 Bradford Assay

Bovine serum albumin (BSA) standards (0.125-4 mg/mL) were prepared in RIPA buffer. 1 µL of standards or cell lysate were diluted into 200 µL of Bradford reagent and vortexed to mix. 50 µL aliquots were placed in a 96 well plate in triplicate. The absorbance was read at 595 nm using Victor^3 1420 multi-label plate reader. A standard curve was calibrated using the absorbance of the BSA standards and sample concentrations calculated from the standard fit.

Bradford Reagent

0.01% (w/v) Coomassie blue G250

4.25% (v/v) Ethanol

8.5% (v/v) Phosphoric acid

2.5.11 Cell growth and invasion assays

2.5.11.1 Scratch assay

A375 FRT- (parental) and FRT-wtAGR2 engineered cells were plated in 100 mm tissue culture plates and allowed to grow to confluency. A wound was simulated using a 10 µL pipette tip to make a straight scratch. Microscopy images were collected at 4 hour intervals and analysed using TScratch software package.

2.5.11.2 xCELLigence cell proliferation assay

Cells to be investigated were trypsinised when 80% confluent (still in logarithmic growth), before counting using a haemocytometer. 5000 cells of each cell type were seeded into wells of an E-plate-view (Roche) in quadruplicate and incubated in cradles of xCELLigence RTCA DP instrument (Roche) at 37^0C, 10% CO_2 for 96 hours. The xCELLigence assay measures cell growth in real time as a measurement of electrical impedance across interdigitated electrodes on the base of the tissue culture plate. The software program was written such that cell impedance readings were taken every 15 minutes over the complete incubation (96 hours and until cells were confluent). Using Real-Time Cell Analyzer software package (Roche) data was analysed, averaged and plotted as changes in growth over time.

2.6 Molecular Biology Techniques

2.6.1 SDS-PAGE separation of proteins

Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) allows separation of complex mixtures of proteins based on their electrophoretic mobility (a function of the length of the polypeptide and its charge) (277;278). SDS is an anionic detergent which linearises proteins and imparts a negative charge to the linearised protein. Thus the proteins will be separated as an approximation of their size. Using the Mini-Protean kit (BioRad), the 0.75 mm resolving gel was first prepared, as per Table 2-7, and overlaid with water to maintain a clean, straight

interface on the surface of the gel. Following polymerisation, the water was removed and the stacking gel was overlaid, directly on top of the resolving gel, and a 10- or 15-well comb added. Once the stacking gel has polymerised, the combs were removed, the gels loaded into the Mini-Protean tank and filled with ~ 1 L of 1× running buffer. Normalised protein samples were prepared (20 µg of protein unless otherwise stated) was concurrently mixed with 4× sample buffer to a final 1× concentration and denatured by heating at 95°C for 5 minutes. Proteins were then loaded to individual wells in the gel, along with pre-stained protein standards (Fermantas) as protein size markers. Electrophoresis was carried out at 200V for ~50 minutes, or until the bromophenol blue dye front of the sample buffer reached the bottom of the gel.

Table 2-7 Final concentrations of reagents required for resolving and stacking gels for SDS-PAGE separation of proteins.

Solution	*12% Polyacrylamide resolving gel*	*13.5% Polyacrylamide resolving gel*	*Stacking gel*
30% Acrylamide mix (30% acrylamide, 0.8% bis-acrylamide)	12 %	13.5 %	5%
1.5M Tris pH 8.8	0.39 M	0.39 M	N/A
1.0 M Tris pH 6.8	N/A	N/A	0.13M
10% (w/v) SDS	0.1%	0.1%	0.1%
10% (w/v) APS	0.1%	0.1%	0.1%
TEMED	0.04%	0.04%	0.1%
Water	As required		

1× Running Buffer

192 mM Glycine

25 mM Tris

0.1% (w/v) SDS

4× Sample Buffer

4 % (w/v) SDS

200 mM Tris-HCl pH6.8

20% (v/v) Glycerol

10 mM EDTA pH 8.0

Bromophenol blue

0.4M DTT (added immediately prior to use).

2.6.2 Protein staining in gels

Following SDS-PAGE, separated proteins could visualised in a number of ways. Firstly, Coomassie Blue staining (see below) was used, whereby the gel is incubated in Coomassie stain for 30 minutes, with constant agitation. Once bands are evident, the gel was destained overnight in destain solution, followed by storage in deionised water. Alternatively, proteins could be rapidly stained using InstantBlue stain (Expedeon), which required only 15 minute incubation with the staining reagent for visualisation of protein bands. When greater sensitivity was required, gels were stained using a silver staining protocol (SilverQuest, Life Technologies) as per manufacturer's instruction. Gels could then be dried onto chromatography paper using a heated vacuum drier (Gel Master Model 1426, Welch Rietschle Thomas).

Coomassie Stain	*Destain*
5% (v/v) Coomassie blue R-450 (Sigma)	7.5% (v/v) methanol
50% (v/v) methanol	10% (v/v) acetic acid
10% (v/v) acetic acid	

2.6.3 Immunoblotting

Immunoblotting for specific protein of interest could be carried out following SDS-PAGE if required. Proteins were transferred from the polyacrylamide gel onto 0.2 µm Hybond-C nitrocellulose membranes (GE Healthcare) in 1× transfer buffer, cooled with an ice pack at a constant current of 300 mA for 60 minutes. The membranes were then ink, or Ponceau S (Sigma), stained to confirm transfer of protein and to assess loading. Prior to immunoblotting, non-specific binding sites were blocked using 3% (w/v) dried skimmed milk powder (Marvel) in PBS-0.05% Tween (PBST), for one hour at room temperature with constant agitation. Subsequently, membranes were probed with primary antibody, at appropriate concentration in 3% milk-PBST, for one hour at room temperature, or overnight at 4°C. The probed membranes were then washed thrice with PBST to remove weak

and non-specific interactions, prior to incubation with a horseradish peroxidase (HRP) conjugated secondary antibody (DakoCytomation) for one hour at room temperature. Membranes were again thrice washed with PBST and prepared on a glass plate for ECL detection. The bound secondary antibody signal was detected using a 1:1 mixture of ECL1:ECL2 solutions, mixed and incubated for 30 seconds before application to the membrane. The membrane was exposed to X-ray film then developed using a Konica Medical Film Processor. Antibodies used are listed in Table 2-8.

1× Transfer buffer

0.192M Glycine

25 mM Tris

20% (v/v) Methanol

ECL Solution I	*ECL Solution II*
100 mM Tris, pH 8.5	100 mM Tris, pH 8.5
2.5 mM Luminol	0.02% (v/v) H_2O_2
0.4 mM p-Coumaric acid	

Table 2-8 Primary antibodies used in this study

Antibody Target	*Molecular Weight (kDa)*	*Description*	*Supplier*	*Dilution*
AGR2 (K47)	19	Rabbit Polyclonal	Moravian Biotechnology	1:1000
AGR2	19	Mouse Monoclonal	Abnova	1:1000
p53 (DO-1)	53	Mouse Monoclonal	Moravian Biotechnology	1:5000
p21	21	Mouse Monoclonal	Merck/Calbiochem	1:500 – 1:1000
β-actin	42	Mouse Monoclonal	Sigma	1:5000 – 1:10 000
IRF-1	48	Mouse Monoclonal	BD Biosciences	1:1000
CHOP (GADD 153)	19	Mouse Monoclonal	Santa Cruz	1:1000
TSG101	44	Mouse Monoclonal	Sigma	1:1000
AGR3 (MAB3.1)	19	Mouse Monoclonal	Moravian Biotechnology	1:1000
AGR3 (MAB3.2)	19	Mouse Monoclonal	Moravian Biotechnology	1:1000
AGR3 (MAB3.3)	19	Mouse Monoclonal	Moravian Biotechnology	1:1000
AGR3/AGR2 (MAB3.4)	19	Mouse Monoclonal	Moravian Biotechnology	1:1000

2.6.4 Confocal immunofluorescence microscopy

Cells were plated on coverslips at 2×10^5 cells per well (40% confluency) in the base of 12-well tissue culture dishes, and incubated at 37°C for 16 hours. Following

incubation, media was removed from wells and cells rinsed with 1 mL of Tris-buffered saline (TBS), followed by fixation in Fixing solution for 10 minutes at room temperature. After fixing, the formaldehyde containing solution was removed, coverslips were retrieved from wells and cells washed twice with TBS-0.1% Triton X-100 (TBS-Triton), prior to permeabilisation and blocking for 1 hour at room temperature with 3% BSA in TBS-Triton. Cells were incubated in a humid environment overnight at 4°C with primary antibodies diluted in blocking solution (1:500 anti-AGR2, Abnova and 1:250 anti-PDI, Enzo Life Sciences). Coverslips were subsequently washed 3 × 5 minutes in TBS-Triton and incubated with appropriate secondary antibodies (Alexa Flurophore conjugated goat anti-mouse 488, or Alexa Fluorophore conjugated donkey anti-rabbit 594) diluted 1:200 in blocking buffer for 30 minutes at room temperature in dark conditions. Cells were subjected to a further 3 washes with TBS-Triton, prior to nuclear counter stain with 100 µL of 1:1000 TO-PRO-3 nuclear stain diluted in TBS. These coverslips were washed a further thrice with TBS-Triton before mounting on glass slides (Superfrost) with mounting fluid for fluorescent analysis (DakoCytomation). Cells were analysed sing an Olympus FV1000 confocal microscope with a 60 × oil immersion objective and FV10-ASW software (Olympus).

Immunofluorescence fixing solution

3.7% Formaldehyde

10 mM EGTA pH8.0

100 mM Pipes pH 6.8

1 mM $MgCl_2$

0.2 % (v/v) Triton X-100

2.6.5 Detection of secreted proteins in tissue culture media

Cells were plated at constant number across experimental conditions, and allowed to grow to confluency over 72 hours. Cells were washed twice with PBS and incubated for 24 hours in serum free media. Media was subsequently concentrated 50× using 10 kDa molecular weight cut off spin column filters (Millipore). Protein concentration of media was quantified by NanoDrop, and standardised (if necessary) to 0.5 mg/mL, prior to boiling with SDS sample buffer, loading onto an SDS-PAGE gel. Gels could be silver stained for whole secretome visualisation or immunoblotted for the protein of interest. Alternatively ^{35}S-methionine labelled proteins were detected using a phosphoimager.

2.6.6 FACS analysis of protein expression

Fluorescence activated cell sorting (FACS) was used to quantify the proportion of cells expressing a protein of interest. Experimental cells were trypsinised, counted using a haemocytometer and diluted to 5×10^6 cells/mL in ice cold PBS supplemented with 10% FBS and 1% NaN_3. 100 µL of each dilution was required per experimental staining condition. Cells were pelleted and fixed in 100 µL of 0.01% formaldehyde for 15 minutes at room temperature prior to permeabilisation. Cells were permeabilised by incubation in 100 µL PBS supplemented with 0.1% Triton X-100 for a further 15 minutes at room temperature. The FITC-conjugated antibody (Anti-Ki67, eBioscience) was prepared in a series dilution (1:20, 1:50 and 1:100), along with negative controls, in the permeabilisation buffer. Following permeabilisation, Cells were incubated with antibody reagent for 30 minutes, in darkened environment at room temperature to allow interaction. After staining, cells were washed thrice by centrifugation at 400g for 5 minutes and resuspension in ice cold PBS with 3% BSA and 1% NaN_3 prior to analysis. Data from 1000-2000 cell events were detected using a FACSAriaII flow cytometer (BD Biosciences), with FITC detected in the green fluorescence channel, and data analysed using FACSDiva software (BD Biosciences).

2.6.7 Quantitative real-time PCR (qPCR)

2.6.7.1 RNA extraction

Cells were seeded in 6-well plates and allowed to grow to near confluency. Cells were harvested as 2.5.8, prior to RNA extraction using RNeasy kit (Qiagen) in accordance with manufacturer's instructions. Samples were homogenised using QIAshredder (Qiagen) columns and RNA eluted using RNase free water. During the procedure careful measures were taken to avoid contamination of RNA. Once isolated, RNA was used immediately to synthesise cDNA.

2.6.7.2 cDNA synthesis by reverse transcription PCR

0.5 μg of RNA was diluted in 7 μL of RNase free water and incubated at 65°C for 5 minutes. RNA was then cooled and combined with 13 μL of the reverse transcription (RT) mastermix. The mixture was incubated for 1 hour at 37°C. Following cDNA synthesis, cDNA could be stably stored at -20°C

Reverse-transcription mastermix

2 μL RT Buffer

2 μL 5 mM dNTP

0.2 μL 0.5 mg/mL oligo DT primers

0.25 μL 40U/μL RNase inhibitor

2 μL 100 mM DTT

1 μL Omniscript Reverse transcriptase (Qiagen)

5.6 μL nuclease free water.

2.6.7.3 qPCR Reaction

qPCR was carried out using reagents from the Solaris kit (Dharmacon). The reaction mixture was prepared as indicated in white qPCR tubes with clear optical caps. Each condition was carried out in quadruplicate. The reaction was monitored on the PTC-200 Peltier Thermal cycler using the indicated cycling program. Once complete, Ct (cycle threshold, the number of cycles to reach an empirical reading of the label) values were obtained, and relative change calculated using the $2^{-\Delta\Delta Ct}$ method (279). ΔCt calculates the difference between the average Ct for the gene of interest and the control gene (e.g. GAPDH). ΔΔCt indicates the difference between ΔCt for the experimental condition compared to control condition.

qPCR Reaction Mix	*Thermocycling Program*
12.5 µL Solaris Mastermix	1. 95°C for 15 minutes
1.25 µL Solaris Primer/Probe set	2. 95°C for 15 seconds
500 ng cDNA	3. 60°C for 1 minute – read plate
Adjusted to final volume of 25 µL with water	Cycle to step 2-3 40 times

2.7 Gene expression analysis by microarray

RNA was extracted, as previously, using the RNeasy RNA extraction kit coupled to the QIAshredder. RNA quality and quantity was assessed with NanoDrop and RNA integrity number (RIN) calculated using Agilent 2100 Bioanalyzer (Agilent Technologies). Purified RNA (0.5 µg) was biotin labelled using the Illumina TotalPrep RNA amplification kit (Ambion) according to the manufacturer's instructions. Labelled cRNA was hybridised to Illumina HT-12 v3 BeadChips and scanned at the Wellcome Trust Clinical Research Facility, Edinburgh. Gene expression analysis was performed using the open source statistical programming language, R (280), and associated Bioconductor packages (281). Probe bead

summary data was quantile-normalised using the Lumi package (282). Differential expression was determined using Rank Products analysis (283) with 5% false discovery rate. Clustering and heat maps were performed and generated using the Cluster and TreeView programs (284). The online Database for Annotation, Visualisation and Integrated Discovery (DAVID) program (285) was used to identify over-represented KEGG (Kyoto Encyclopaedia of Genes and Genomes) pathways from lists of differentially expressed genes. Statistical analysis was carried out in collaboration with Dr A. Sims, Applied Bioinformatics of Cancer, Edinburgh Cancer Research Centre.

2.8 Expression and Purification of Recombinant protein from *E. coli*

2.8.1 AGR2 protein expression from *E. coli*

Hexa-histidine tag protein expression vectors (pEHISTEV (286)) cloned with wild-type or mutant $AGR2_{21\text{-}175}$ or AGR3 genes were heatshock transformed into *E. coli* BL21-DE3 cells and selected on LB-agar supplemented with 50 µg/mL kanamycin. A single colony was isolated, and inoculated into 50 mL of LB (+ 50 µg/mL kanamycin) and incubated at 37°C with constant shaking (220 rpm) overnight. This starter culture was diluted 1:20 into 500 mL LB containing selective antibiotic, and returned to the incubator until an $OD_{600\ nm}$ of 0.6 was achieved. Protein expression was subsequently induced with the addition of 1 mM IPTG. The culture was incubated for a further 3 hours at 37°C with constant shaking. Following induction, cells were centrifuged for 20 minutes at 6000 rpm at 4°C to pellet. Pellets could be snap frozen and stored at -80°.

2.8.2 Purification of His-tagged AGR2

Two methods of His-tag purification were carried out. For small scale purifications, bacterial pellets were resuspended in 10 mL of Lysis buffer and incubated on ice for 30 minutes. Cells were further lysed by sonication, in an ice bath, in three bursts of 15 seconds with 30 seconds gap between steps. Lysate was then cleared of insoluble material by centrifugation at maximum speed for 30 minutes at 4°C. The supernatant was transferred to a tube containing 1 mL of Ni^{2+}-NTA agarose beads (Qiagen), which had been pre-washed with 5 mL of lysis buffer. The supernatant and beads were incubated at 4°C on a rotating table for 1 hour. The mixture was transferred to a disposable purification column and allowed to flow through by gravitation force. The beads retained in the column were washed 2× 5 mL with lysis buffer and 2× 5 mL with Wash buffer. Following washes, bound His-tagged protein was eluted in 15 elution steps of 500 μL of high imidazole Elution buffer. Fractions were collected and analysed by SDS-PAGE and colloidal staining, and protein containing fractions pooled for immediate use or storage at -80°C. To remove the His-tag leaving a 4 amino acid overhang, protein was prepared in buffer containing 0.1M DTT and 330 U AcTEV enzyme (Life Technologies). A time course (1-24 hours) of recombinant tobacco etch virus (TEV) mediated cleavage was carried out to assess the rate of cleavage. The digested protein was dialysed against Lysis buffer, or buffer exchanged using Zeba desalting columns (Thermo Scientific), to reduce high imidazole content before being purified over nickel-agarose beads to remove cleaved His-tag, and His-tagged protease.

Lysis Buffer	*Wash Buffer*	*Elution Buffer*
50 mM NaH_2PO_4	As Lysis Buffer except with 30 mM imidazole	As Lysis buffer except with 300 mM imidazole
400 mM NaCl		
10 mM imidazole		
1 × protease inhibitor cocktail		

Alternatively, for larger scale purifications, bacterial pellets were resuspended in 30 mL of Lysis buffer on ice and vortexed. Lysis was achieved by high pressure cell disruption at 25k PSI using a 1.1 kW TS cell disruptor (Constant Systems), followed by centrifugation at 40,000 × g for 50 minutes to isolate the soluble fraction. The supernatant was applied to a nickel-charged HisTrap FF 5-mL column (GE Life Sciences), pre-equilibrated with lysis buffer in an AKTA chilling cabinet (Edinburgh Protein Production Facility). The immobilised protein was washed with 20 column volumes of wash buffer, followed by 10 column volumes of wash buffer supplemented with 30 mM imidazole, before final elution using 300 mM imidazole for 20 column volumes. The eluted protein was pooled and cleaved with TEV protease overnight at 20°C. The resultant solution was desalted into lysis buffer with a HiPrep desalt 26/10 column (GE Life Science) before reloading onto the IMAC to remove protease and uncleaved protein. The protein was collected as flow through and protein concentrations determined by NanoVue (GE Life Sciences).

2.8.3 Size-exclusion chromatography of purified AGR2

Purified protein was analysed by gel-filtration size exclusion chromatography on an analytical Superdex PC75 3.2/30 column with a flow rate of 0.5 mL/minute at titrated concentrations ranging from 2.5 mg/mL to 5 μg/mL. Detection was by absorption at 214 nm. The column was previously calibrated with known protein standards (GE Life Sciences).

2.8.4 Purification of GST-tagged Reptin

Reptin-GST tagged vectors for protein expression (a kind gift from Dr M. Maslon) were transformed into BL21-AI cells. These were cultured as 2.8.1, but induced with 0.2% arabinose at room temperature for 3 hours. Following induction cells were pelleted as previously. The cell pellet was resuspended in 10 mL of Lysis buffer and incubated on ice for 30 minutes. Cells were then further lysed by sonication in 3 × 15 second bursts. The lysate was centrifuged for 30 minutes at 4°C at maximum speed. The supernatant was transferred to a fresh tube containing 500 μL of glutathione-sepharose 4B beads (Amersham), pre-equilibrated with 5 mL lysis buffer. Beads and lysate were incubated for 1 hour at 4°C with constant agitation. The beads were subsequently washed twice with 5 mL high salt wash buffer, twice with 5 mL low salt wash buffer, prior to the addition of 5 mL of elution buffer. Elution buffer was incubated with the beads for 30 minutes at 4°C, and eluted protein collected as flow through.

Lysis Buffer

10% Sucrose

50 mM Tris pH 8.0

400 mM NaCl

0.5 mg/mL Lysosyme

0.5% Triton X-100

1 mM DTT

1 mM benzamidine

1 × protease inhibitor cocktail

High Salt Wash

20 mM Hepes pH 7.5

1 M NaCl

1 mM DTT

1 mM benzamidine

Low Salt Wash

20 mM Hepes pH 7.5

50 mM NaCl

1 mM DTT

1 mM benzamidine

Elution Buffer

100 mM Tris pH 8.0

120 mM NaCl

40 mM reduced glutathione.

2.9 Quantitation of protein expression levels using SILAC mass spectrometry.

2.9.1 Metabolic labelling of cellular proteins and preparation of cell lysate

Experimental engineered A375-FRT cells containing the required gene recombination were incubated with isotope labelling media as described in the figure legends. Briefly, media (Dundee Cell Products) containing normal isotope labelled lysine and arginine amino acids (R0K0, 'light'), or heavy isotopes, ^{13}C labelled Arg and ^{2}H labelled Lys (R6K4, 'medium') or ^{13}C and ^{15}N labelled Arg and ^{13}C and ^{15}N labelled Lys (R10K8, 'heavy'), supplemented with 10% dialysed FBS (Dundee Cell Products) was prepared. Cells were propagated for >6 cell doublings to ensure incorporation of the isotope labelled amino acids (287). A near-confluent (95%) 100 mm tissue culture plate of the desired experimental condition was harvested and lysed with RIPA cell lysis buffer, as 2.5.9. Protein concentration was determined by Bradford assay and samples mixed in a 1:1:1 ratio. The mixed sample reduced in SDS-PAGE loading buffer containing 10 mM DTT, and alkylated in 50 mM iodoacetamide prior to boiling and separation by one-dimensional SDS-PAGE (using pre-cast 4-12% Bis-Tris Novex mini-gel, Life Technologies), and visualised by colloidal Coomassie staining. The entire protein gel lane was excised and cut into 10 gel slices. Each gel slice was subject to in-gel digestion with trypsin (288). The resultant tryptic peptides were extracted with 1% formic acid, lyophilised in a Speedvac (Helena Biosciences), and resuspended in 1% formic acid.

2.9.2 Mass spectrometry methods

Trypsinised peptides were separated using an Ultimate U3000 (Dionex Corporation) nanoflow liquid chromatography (LC)-system. 2 mg of digested peptides were loaded with a constant flow of 20 mL/minute onto a PepMap C18 trap column (0.3 mm id × 5 mm, Dionex Corporation) for reverse phase chromatography. After trap enrichment, peptides were eluted onto a PepMap C18 nano column (75 mm × 150

mm) with a linear gradient of 5-35% solvent B (90% acetonitrile, 0.1% formic acid). The HPLC system was coupled to an LTQ Orbitrap velos (Thermo Scientific) via a nano ES ion source (Proxeon Biosystems). Full scan MS survey spectra (m/z 335-1800) in profile mode were acquired in the Orbitrap, with a resolution of 60, 000 after accumulation of 500,000 ions. The five most intense peptides ions from the preview scan were fragmented by collision-induced dissociation (CID) for MS/MS sequencing. Data were acquired using the Xcalibur software (Thermo Scientific). Mass spectrometric analysis was carried out by Dundee Cell Proteomics, Dundee.

2.9.3 Relative Quantification and Bioinformatics analysis

Quantification was performed using MaxQuant version 1.0.7.4 (289), by Dundee Cell Proteomics. To minimise the effect of outliers, protein ratios were calculated as the median of all SILAC pair ratios that belonged to peptides contained in the protein. The generation of peak list, SILAC quantitation, calculated posterior error probability (PEP), false discovery rate, peptide to protein group assembly, and data filtration was carried out using MaxQuant. The derived peak list was searched with Mascot search engine (Matrix Science) against a concatenated database from International Protein Index (IPI) human protein database version 3.6 (forward database), and reversed sequences of all proteins. Parameters allowed up to three missed trypsin cleavages and three labelled arginine and lysine amino acid residues. Initial mass deviation of precursor ion and fragment ions were up to 7 ppm and 0.5 Da, respectively. The minimum required peptide length was 6 amino acids. For statistical evaluation, PEP for peptide identification should be below or equal to 0.1, and the false positive rate was set to 5% at the peptide level. Proteins were quantified based on the presence if at least one MaxQuant-quantifiable SILAC pair was evident. Enzyme specificity was set to trypsin allowing for cleavage of N-terminal to proline and between aspartic acid and proline. Carbamidomethylation of cysteine was searched as a fixed modification, and N-acetyl protein and oxidation of methionine were searched as variable modifications.

Data was mined as described in results and figure legends. SILAC quantified relative expression ratios requiring a minimum fold change of 20% were empirically defined. Expression changes ±20% with a minimum of 3 peptides used for quantitation were identified and submitted to Ingenuity Pathway Analysis (Ingenuity, Dr. J Nicholson, Australian Proteomics Facility, Sydney), allowing the identification of over-represented factors and comparison to a curated library of expression changes on biochemical function. In addition, data-driven analysis was carried out through comparison of proteins of greatest change in relative expression levels with literature review, and thorough biochemical validation of results. Biochemical validation required immunoblot studies, coupled with gene over-expression and silencing of specified gene of interest.

2.10 Assays

2.10.1 Peptide ELISA

White 96-well well microtitre plates (Costar) were coated with 1 μg per well in 50 μL PBS streptavidin (Cambridge Bioscience) and incubated at 37°C overnight. Wells were subsequently washed 4× 200 μL with PBST, to remove excess streptavidin protein. 0.5 μg of biotinylated peptide per well in 50 μL of water was added and incubated for one hour at room temperature on a rotating table. Following peptide addition, the plate was washed 6× with 200 μL PBST, before blocking with 200 μL of 3% BSA in PBST and incubated for one hour at room temperature with shaking. Subsequently the protein of interest, as indicated in the figure legend, was diluted as required in 50 μL of 3% BSA in PBST and added to the wells for one hour at room temperature. The wells were washed 6× with PBST, as previously, and 1:1000 dilution of the appropriate primary antibody in 50 μL of 3% BSA in PBST was applied to the wells and incubated for a further one hour at room temperature. Following a further 6 washes with PBST, before incubation with the appropriate HRP conjugated secondary antibody (diluted 1:1000 in 50 μL of 3% BSA in PBST) and incubated for 1 hour at room temperature. A final 6 washes with PBST were

carried out, ECL reagents were mixed, and 50 µL applied to the well. Luminescence was read using a plate reader (Fluoroskan Accent). All experiments were carried out at least in triplicate, and representative of 2 experiments

For epitope mapping experiments, diluted antibody of interest (in 3% BSA in PBST) was applied immediately following peptide incubation and subsequent washes. Incubation was carried out for one hour at room temperature before detection using secondary antibody and ECL as previously.

2.10.2 Protein ELISA

96-well microtitre plate was coated with 100 ng of protein (as per legend) in 50 µL of ELISA coating buffer (100 mM $NaHCO_3$, pH 8.6) at 4°C overnight. Subsequently, wells were washed with 6× 200 µL of PBST and non-specific interaction sites blocked by one hour incubation with 3% BSA in PBST on a shaking table. A titration of the interacting protein (as per legend) was prepared with a final volume of 50 µL in 3% BSA in PBST, and applied to well. Protein incubation was carried out for one hour at room temperature with gentle agitation. Following this, wells were washed with 6 × 200 µL PBST. An appropriate primary antibody was added to wells (1:1000 dilution in 3% BSA in PBST), and incubated for one hour at room temperature. Another round of 6 washes with PBST was carried out, and a 1:1000 dilution of HRP-conjugated secondary antibody (50 µL) applied. After a final 6 washes with PBST, interaction was detected and quantified with ECL and luminescence detection. All experiments were carried out at least in triplicate, and representative of at least 2 experiments

2.10.3 Two-site sandwich microtitre assay (^{2S}MTA)

2.10.3.1 DyLight800 Fluorophore labelling of antibody

MAB3.4 monoclonal antibody was conjugated to DyLight800 (DyL800) fluorophore (excitation at 770 nm and emission at 794 mm) using DyLight800 Microscale Antibody Labelling kit (Thermo Scientific) according to manufacturer's instructions.

Briefly, 50 µL of MAB3.4 (1.3 mg/mL) was diluted to a final volume of 100 µL with 5 mM Borate buffer and the protein added to a vial of DyLight800 reagent to initiate coupling and the reaction incubated in darkness for 60 minutes at room temperature. The labelled protein was applied to a spin column in microfuge format, and the labelled antibody separated from unreacted label by collecting the eluate after centrifugation. The solution of MAB3.4-DyL800 was stored at 4°C for up to four weeks.

2.10.3.2 ^{2S}MTA detection of AGR2 protein oligomerisation

Capture immunoglobulins (unlabelled MAB3.4, 100 ng/50 µL) in the solid phase were adsorbed onto 96-well black walled, clear bottomed 96-well plates (Costar), diluted in 100 mM $NaHCO_3$, pH 8.6 and incubated at 4°C, with agitation for 14-16 hours. Wells were subsequently washed 3-5 times with PBST, prior to non-specific site blocking with 200 µL of 3% BSA in PBST per well for 1 hour at room temperature. Concurrently 100 ng (5.5 pmoles) of purified AGR2 protein was diluted to 50 µL per well in 3% BSA in PBST and incubated with the relevant screening condition (overlapping peptides of AGR2 or Strathclyde Institute for Drug Research (SIDR) library compounds) for one hour. The blocking solution was removed from the plate, and replaced with the protein-condition mix solution and incubated on the plate for 1 hour at room temperature. Protein was then aspirated and the plate washed 3-5× with PBST prior to antibody detection. Detection antibodies, MAB3.4-DyL800 were diluted to 150 ng per well in 3% BSA in PBST and 50 µL per well incubated on the wells for 1 hour at room temperature. Plates were washed a final 3-5× with PBST prior to detection. DyL800 fluorophore was directly detected on the Licor Odyssey with 750 nm laser excitation and emission at 800 nm. Images were analysed and quantified with the Odyssey Sa software (Licor).

Alternatively, to measure the oligomerisation status of mutant AGR2 proteins, 500 mL pellets of bacteria induced to express the protein of interest, prior to lysis and sonication. The lysed bacteria were centrifuged to isolate the soluble fraction. 10 µL fractions of each mutant crude lysate was separated by SDS-PAGE,

electrophoretically transferred to nitrocellulose membrane and probed with the DyL800-labelled MAB3.4 antibody. Using the Licor Odyssey system, an immunoblot was used to relatively quantify and normalise AGR2 protein concentrations. The ^{2S}MTA was carried out as previously, and normalised AGR2 protein applied in the mobile phase, prior to DyL800-MAB3.4 and near-IR detection.

2.10.4 Oligomer cross-linking using DSS

50 µL of normalised AGR2 induced crude bacterial lysates, or mutants thereof, were incubated with titrations of Disuccinimidyl suberate (DSS) from 5-1000 µM for one hour at room temperature. The cross linking reaction was quenched following this incubation by the addition of 2.5 µL 1 M Tris pH 8.0. Next, SDS-PAGE sample buffer was added, and samples heated for 5 minutes at 95°C, prior to SDS-PAGE and subsequent immunoblotting to detect quaternary structure.

2.11 Protein crystallisation methods

Screens were prepared in 24-well or 96-well plates (Linbro). Mother liquor well solutions were empirically defined, or pre-prepared from commercial crystallisation screens as described (Molecular Dimensions). Purified protein was concentrated using spin columns with a molecular weight cut off of 5 kDa, to the required concentration. Droplets were prepared consisting of 50% purified protein and 50% well solution either by hand or using the Oryx 8 robot (Douglas Instruments). Crystal presence and phase separation were first analysed by the operator, or alternatively for high throughput using CrysCam X-Y stage coupled with CrysScore software (Art Robbins Instruments). Crystallisation conditions were incubated at 18°C. Seeding solutions could be used to initiate protein nucleation, using a condition which shows small protein micro-crystals diluted 1:10 in protein buffer than supplemented to the drop at 10% of final volume prior to incubation. Following

incubation, crystals were dipped in a 40% PEG freezing solution prior to immersion in liquid nitrogen. Crystals were exposed to X-ray at Diamond Light Source.

Chapter 3: Generation and Characterisation of an Isogenic Stable Cell Panel for the Investigation of AGR2 Function

3.1 Introduction

The efforts of the Human Genome project, and other studies to map the genome sequence of other organisms, have provided us with data of the existence of tens of thousands of genes and their potential expressed products. The human genome is estimated to consist of 20,000 – 25,000 unique genes (290) rather than the ~100,000 originally estimated (291). The biological machinery of the cell is diversified by the numerous post-translational modifications, secondary promoters, alternative splice sites and differential transcription termination, resulting in an estimated total number of proteins to be over one million (292-294). Despite this plethora of information, the understanding of essential processes is staggeringly limited, such that thousands of known proteins remain functionally lacking description, almost ten years since their first elucidation. Understanding the specific function of a protein in a signalling system might provide an insight into a physiological process, and the understanding of therapeutic strategies for a protein prognostic of a disease outcome. The network of regulatory systems, and resultant signalling cascades, mediated through protein-protein interactions can provide a comprehensive picture of derived protein function.

Previously described functions of Anterior Gradient-2 include roles in development (149;151;159;207;244;295), endoplasmic reticulum stress response (165;179), protein maturation (165;233) and cell growth (150;164;244;249;270), coupled to this; complexities in the sub-cellular distribution of the gene product (187;188), present an intricate network of interplay which describes the diversity of AGR2-associtated pathways. The identification of AGR2 as an over-expressed feature in a vast range of human cancers, its discovery as an inhibitor of p53 (150), and the perturbation of the EGF receptor cascades (194) has driven efforts to deconvolute the

signal transduction pathways influencing pro-oncogenic functions to assess AGR2 as a possible therapeutic target for novel anti-cancer drugs. Upstream induction by chemical agents promoting endoplasmic reticulum stress has been thoroughly characterised (179). Additionally, the transcriptional induction of the *agr2* gene by oestrogen (175;176) and anti-oestrogen drug compound, tamoxifen (174;272), as well as serum independent signals (178) have been well described. Conversely, *agr2* suppression by the SMAD4-TGFβ signalling axis highlight an oestrogen independence to cell signalling (180). These build on the fields understanding of AGR2 induction and purpose in the stress response, while addressing the subsequent molecular effect of this induction remains elusive.

The majority of studies interrogating downstream AGR2 effects have focussed on the direct interaction between AGR2 and protein interaction partners, to build an interactomic network which describes the interplay between molecules resulting in the global AGR2-dependent output (Table 1-2). For example, the identification of AGR2 co-immunoprecipitation with MUC2 (165) has been used to characterise the function of AGR2 as an ER chaperone protein, assisting in the maturation of cysteine rich nascent peptide derivatives. The yeast-2-hybrid approach has identified C4.4A and α-dystroglycan (172), as potential co-factors involved in the metastatic nature associated with AGR2; however these interactions have not been thoroughly biologically validated as *bona fide* protein-protein interaction partners in human cells (199). Due to the false positive prone nature of yeast-2-hybrid studies, thorough validation of partners is necessary to demonstrate biological significance (296). Subsequently, the ATP binding protein Reptin has been identified by yeast-2-hybrid followed by thorough validation providing the first comprehensively confirmed AGR2 binding protein (199); such that the interface of AGR2 binding to Reptin has been mapped to a specific domain, amino acid residues 104-110, representing a divergent loop shared with ERP18. Accordingly, AGR2 is thought to have some regulatory effects on the many functions of Reptin (ATPase activity, ATP binding, helicase functions, telomerase/Pontin binding, APPL1/2 binding, TIP60 interactions among others). Ongoing cellular studies will determine whether Reptin forms a key determinant in the cell signalling pathways modulated by AGR2 expression.

An alternative approach to defining the AGR2 interactome sought the use of a combinatorial phage-peptide library to search for a high affinity peptide motif binding to AGR2 (275). AGR2 was shown to have peptide binding activity by ELISA and complementary Western blot assays to the peptide motif (S/T)xIhh (where x represents any amino acid, and h is an amino acid with a hydrophobic side chain). Proteins containing these motifs could diversify the AGR2 interactome landscape. The (S/T)xIhh peptide expressed in a GFP-fusion vector could increase AGR2 protein levels and induce p53 activity (187). Indeed several proteins have been identified containing these motifs and are subject to ongoing study (Table 3-1. Personal discussion E. Murray, M. Maslon, J. Nicholson, M. Lawrence and T.Hupp). In addition, two groups concurrently utilised a protein crosslinking approach to identify interacting partners of AGR2 in close proximity using cell permeable covalent crosslinking agents, prior to immunoprecipitation and identification by mass spectrometry (202;203). Despite observing high molecular weight cross-linked complexes, the dominant crosslinked partner protein in these studies was identified as AGR2 itself, signifying that AGR2 appeared to form homodimeric complexes *in vivo* and *in vitro*.

Table 3-1 Proteins containing the (S/T)xIhh motif for further study of AGR2 interactome

Protein	Motif	Role
TMEM67/Meckelin	PTPIFY	Transmembrane protein involved in cilia formation
SMG-7	LPTLIYY	Non-sense mediated decay factor
HECTD1	STIFY	E3 ligase
TMEM63B	PTIVYY	Transmembrane protein
HERC2	LTTEFG	E3 ligase

To date few studies have begun to try to disentangle the molecular effect of AGR2 expression in cancer pathways. Dong *et al*. (194) describes AREG as an effector molecule of AGR2, implicating the Hippo signalling pathway co-activator, Yap-1. AGR2 induces the nuclear localisation of Yap-1, thought to be a dephosphorylation effect, and resultant expression of Yap-1 transcriptional targets, including AREG. The resultant elevation of RNA and protein expression of AREG, but not that of other EGFR ligands, is the major stimulatory factor of the EGFR signalling pathway,

and could explain the increased cell proliferation and anchorage-independent growth observed in previous studies (164;218). This insightful study begins to develop the molecular mechanisms whereby AGR2 achieves cellular effects, but does not look broadly at the cell-wide effect of AGR2 up-regulation.

The difficulty of interrogating the downstream effect of AGR2 expression lies in the development of appropriate models. *In vitro* transient transfections and RNAi based methodologies present a rather crude mechanism to measure AGR2 output since there is a lack of acute control of the level of gene expression exogenously induced or silenced. As such, detected outputs may be as a result of stress mechanisms induced in the cells following modified protein synthesis machinery, or stress pathways induced by the transfection reagent itself. In this study, I present a methodology to prepare a stable cell line, constitutively expressing AGR2 for use in a systems biology approach to begin to understand the function of AGR2 overexpression in multiple pathologies. I also present data on the engineering of a stable cell line, and characterisation therein of a novel tool for the interrogation of AGR2 signal transduction.

3.2 Results

3.2.1 Preparation of a cell line incorporating the Flippase recombination target site into the genome

The AGR2-negative A375 human melanoma cell line was chosen as the model system to define the ability of the artificially-recombined AGR2 allele to alter proteostasis and reprogramme a cell into a pro-oncogenic mode. The cell model exhibits a well profiled wild-type p53 pathway responsiveness (297), toll-like receptor 3 (TLR3) interferon system signalling (298), cell cycle checkpoint pathway and DNA damage responsive network that provides a template from which the introduction of the AGR2 gene product can be evaluated through alterations in homeostatic growth control, and link to the p53 pathway. We sought to use a cell

line which stably integrated a single gene allele, using a promoter system that expressed relatively modest levels of the target gene to determine how low level introduction of an onco-protein can alter cell signalling pathways involved in the phenotypically described activities of AGR2.

The initial stage of the Flp-In system requires the integration of a single Flippase recombination target (FRT) site into the genome of the host cell line (Figure 3-1A). A375 cells were first transfected with the pFRT/*lac*Zeo plasmid prior to subculture into single clones for analysis of single site integration. The pFRT/*lac*Zeo contains a single 34 base pair FRT sequence which serves as the target for Flp recombinase (299-301), and a *lac*Z-*Zeocin* fusion gene under the control of an SV40 promoter which codes for Zeocin antibiotic resistance (Figure 3-1B). Successful integration permits the isolation and proliferation of zeocin-resistant foci. Resistant foci were isolated (24 in total), genomic DNA extracted and screened by Southern blot (protocol devised and carried out by Dr. Khaldoon Al-Sammam, University of Edinburgh). Genomic DNA was digested with *Hind III* which digests at a single site within the minimal FRT site (but not within the *lac*Z gene) of the pFRT/*lac*Zeo, before being probed with a ^{32}P radio-labelled probe specific for the *lac*Z gene. Due to the random nature of genetic transfection, the pFRT/*lac*Zeo plasmid may integrate numerous times into the genome of the host cell line, and it is imperative at this stage to identify and discard clones which contain more than one FRT site. Multiple FRT sites may result in increased chromosomal rearrangements or unexpected recombination events as a result of complementary FRT sites being in near proximity. Successful integration is confirmed by the appearance of a single band in the Southern blot, corresponding to the 400 base pairs of the *lac*Z gene. Mammalian cells do not endogenously contain a *lac*Z gene; therefore clones containing pFRT/*lac*Zeo could be easily identified. The Southern blot band varies depending on the location within the genome and the different sizes of the *Hind III* fragments. Selected results (Figure 3-1C) indicate clones with single FRT site integration (clone 1), no integration (clone 2, possible error with restriction digest and/or probe hybridisation) and multiple FRT site integration (clone 3) which demonstrate distinct

location within the genome of the host cell line. In this representation, only clone 1 can be progressed for further analysis.

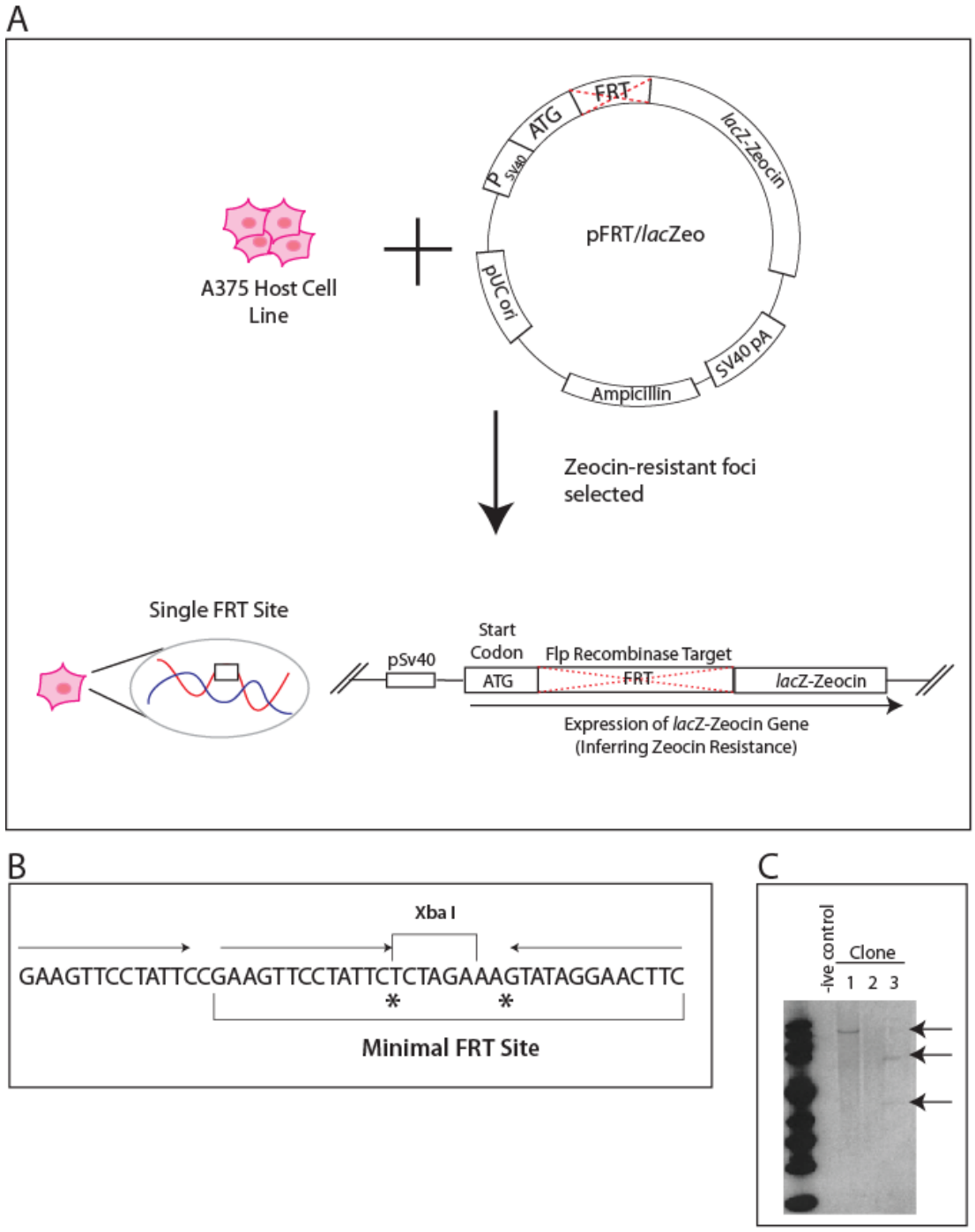


Figure 3-1 Incorporation of the FRT site into the A375 host cell line genome (A) Schematic illustrating the transfection of A375 cells with pFRT/*lac*Zeo plasmid and the integration of the single FRT site into host cell genome. The *lac*Z-Zeocin fusion gene incorporated into the vector allows selection for Zeocin-resistance. (B) The minimal FRT site consisting of 34 bp sequence containing

the two imperfect 13 bp inverted repeats separated by an 8 bp spacer that includes an *Xba I* restriction enzyme digest site. The Flp recombinase recognises this site and mediates recombination between two FRT sites of interacting DNA molecules, the cleavage sites are signified by the two asterisks (*). (C) Southern blot of A375 cells transfected with pFRT/*lac*Zeo and cultured in medium containing 400 μg/mL Zeocin for 14 days. 24 single surviving clones were isolated and repopulated. The DNA of these clones were extracted and digested with *Hind III* (contains a single restriction digest site within the FRT site), followed by analysis using a ^{32}P radioactive labelled probe specific for the *lac*Z gene. Results show 3 representative clones with a single FRT site integration (black arrows), which demonstrate a distinct location within the host cell lines. In some Zeocin resistant cells, no identifiable FRT insertion, as in clone 2, was detected (possible error in digestion and probe hybridisation), and some incorporate more than one FRT site, as clone 3. Only single integrants were progressed.

3.2.2 Validation of the normal functioning signalling pathways of the A375-FRT cell line

Once single site integrants were confirmed, the clonal cell line was tested for normal functioning pathways (Figure 3-2). It had be previously been shown that AGR2 functions to promote cisplatin resistance in xenograft models (248), and the protein can be over expressed in cancers which are targeted for therapy using cisplatin. Host FRT-containing cell lines were challenged as to whether or not they retain the ability to activate the p53 tumour suppressor, and resultant p21 up-regulation in response to cisplatin-induced DNA cross-linking (Figure 3-2A). In addition, the TLR3-responsive interferon system represents a second p53-dependent signal responder; the treatment of cells with poly-inosinic:poly-cytidylic acid (poly I:C) stabilises the IRF-1 protein (Figure 3-2B, data courtesy of Dr K. Al-Samman), indicating that the parental host cell line retains p53 signalling integrity (297;298). Poly I:C is a synthetic analogue of double stranded RNA, allowing simulation of viral infection, This is recognised by the TLR3-receptor pathway and subsequent induction of cytokines (302).

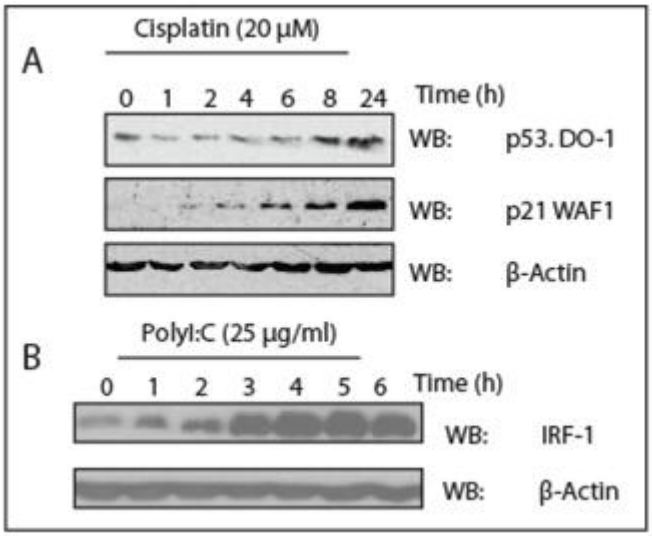


Figure 3-2 Evaluation of the integrity of the A375-FRT parental cells (A) Evaluation of a functional wild-type p53 pathway was defined by the ability of p53 to be activated by incubating cells with 20 µM cisplatin over a 24 hour time course. The induction of p53 protein results in a subsequent induction of p53 activity-dependent gene product p21. Cells were lysed in 1% NP40 lysis buffer and 20 µg of protein lysate was loaded onto a 13.5% polyacrylamide gel and probed for p53, p53-dependent activity using p21-WAF1. (B) Cells also demonstrate an intact IRF-1 response to 25 µg/mL poly I:C treatment over a 6 hour timescale, β-actin is shown as a loading control.

3.2.3 Cloning of the AGR2 gene into the FRT-site containing expression vector and engineering of the constitutively expressing AGR2 experimental cell line.

Concurrently, utilising the Gateway technology (Life Technologies, UK) the gene of interest can be PCR amplified from a template using prepared forward and reverse primers incorporating the attB sites (Figure 3-3). The PCR product is then recombineered, using the BP clonase enzyme, into the empty pDONR221 backbone vector which has the complementary attP sites, prior to being transformed into *E. coli* and selected for kanamycin resistance. This process removes the ccdB/CmR from the pDONR221, allowing the insertion of the gene of interest (GOI), and forms the attL sites important for the second step of the process. Once positive clones were identified, isolated and DNA extracted, the second reaction, the LR reaction, was prepared. This involves incubating the entry clone (pDONR221-containing GOI) with the destination vector, and the LR clonase enzyme which catalyses the recombination of gene of interest via the attL sites with the attR sites of the destination vector, in this case pEF5/FRT/V5-DEST. The pEF5/FRT/V5-DEST vector contains key elements for the expression of the gene of interest. (i) The FRT site in the backbone is identical to that of the pFRT/*lac*Zeo, and subsequently mediates the integration into the host cell (Figure 3.1B); (ii) the bovine growth hormone (BGH) poly-adenylation sequence allows the proper termination and processing of the recombinant plasmid and (iii) the hygromycin resistance gene, which lacks a promoter and an ATG initiation codon. Once the host cell line is confirmed to only integrate one FRT site, and the gene of interest is cloned into the FRT-site containing vector then the reagents for the engineering of a stable cell line are prepared.

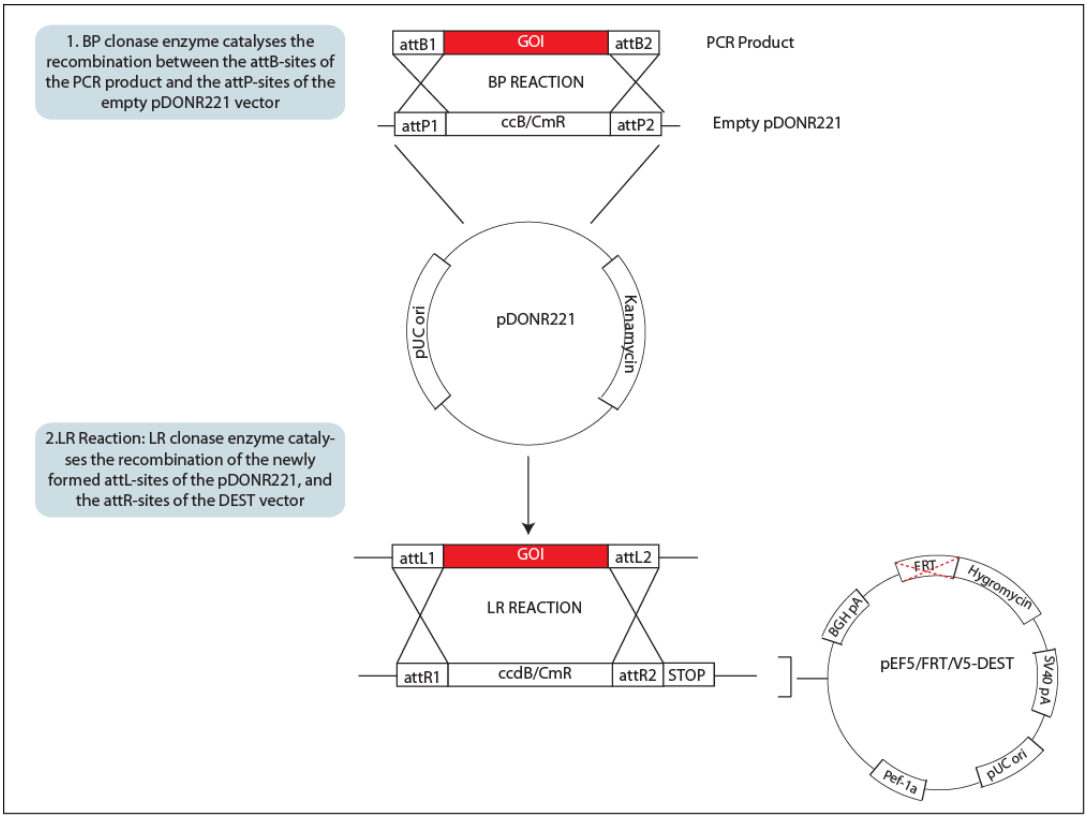


Figure 3-3 Schematic of the Gateway cloning technology allowing the recombination of the gene of interest, in this case AGR2 or AGR3, into the Flippase Recombination Target-containing destination vector, pEF5/FRT/V5-DEST.

The A375-FRT host cell line was co-transfected with pEF5/FRT/V5-DEST containing the gene of interest, and pOG44 in a ratio of 1:9 (Figure 3-4). pOG44 vector expresses the Flp-recombinase enzyme which recognises the FRT sites integrated into the A375 cell genome and the protein expression vector, and catalyses the homologous recombination event. This recombination event incorporates the gene of interest into the host cell genome, while bringing the SV40 promoter and ATG initiation codon into frame with the hygromycin resistance gene and inactivates the *lac*Z-Zeocin fusion gene. The AGR2 gene is also placed under the control of the strong regulatory human cytomegalovirus (CMV) promoter from the pEF5/FRT backbone. This is a key requirement due to the presence of only a single allele of the AGR2, and as a result the expression of the gene product at an appropriate level for biochemical detection. The recombination event is carried out for 24 hours at 32°C, the optimal temperature for the catalysis of the reaction. Subsequently, media was then exchanged and challenged with hygromycin-supplemented media for selection of positively recombineered cells. Following this step, all hygromycin resistant foci are isogenic, containing a single integrated copy of the gene of interest, and all exhibit this gene at the same location in the genome under the control of the same promoter.

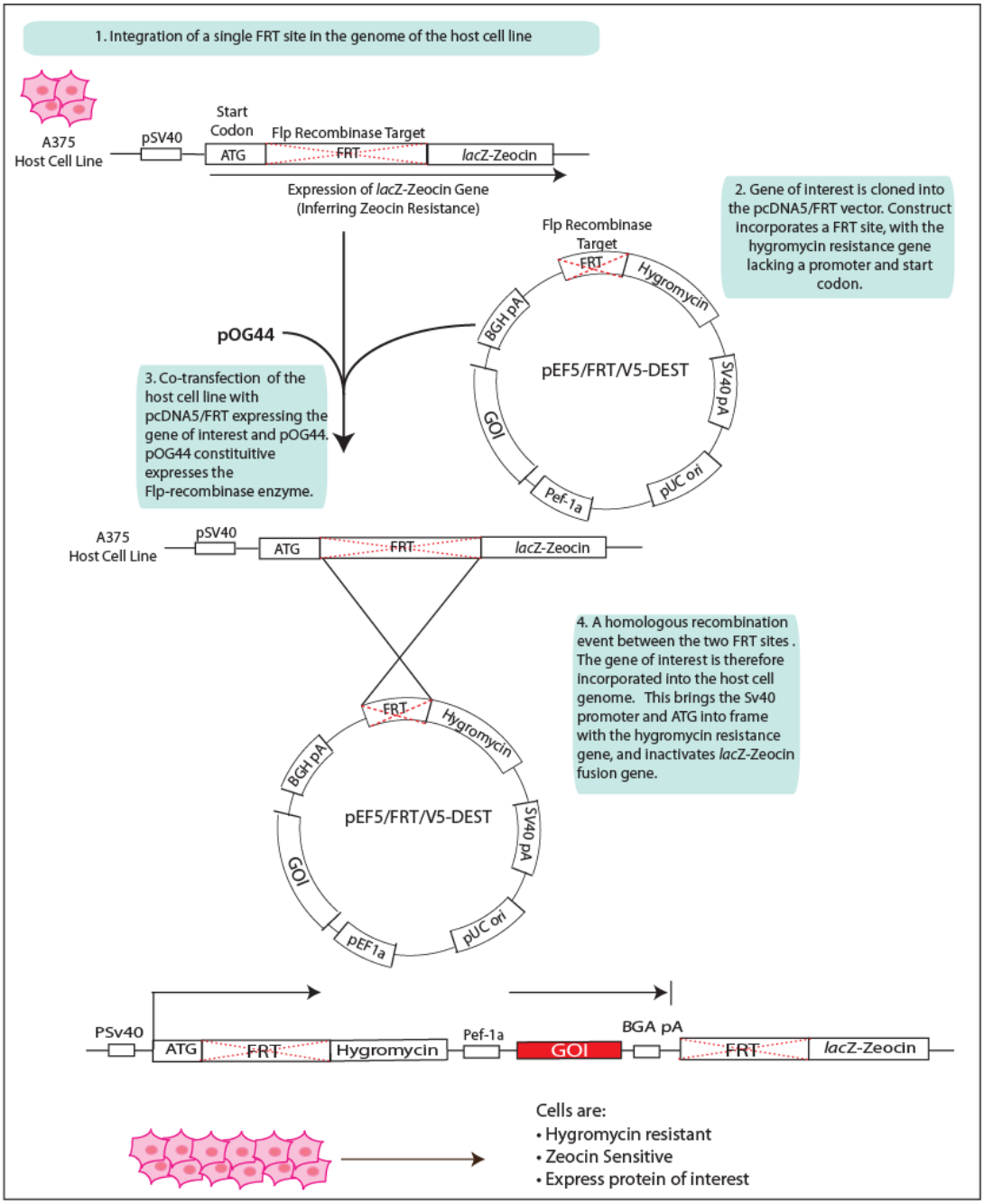


Figure 3-4 Illustration of the engineering of stable cell lines stably expressing a gene of interest utilising Flp-In methodology. A375 cells integrating the single FRT site are co-transfected with pEF5/FRT/V5-DEST expression construct containing the gene of interest and pOG44 plasmid expressing the Flp-recombinase enzyme, as indicated. Following hygromycin antibiotic selection, isolated cells are assayed for the expression of the protein of interest.

3.2.4 Characterisation of the A375-FRT-AGR2 cell line

Following the hygromycin selection of isogenic cells, cell lysates were then assayed for the expression of the protein of interest (Figure 3-5A) and Zeocin-sensitivity (data not shown). As expected, the parental cell line expresses no AGR2 protein, while the cells engineered to express AGR2 exhibits distinct constitutive levels of expression. Expression levels of AGR2 in the A375-FRT-AGR2 cell line were analysed relative to the expression levels of known endogenous AGR2-expressing cell lines (Figure 3-5B). The breast cancer cell lines MCF7 and T47D, oesophageal cancer line OE19 and lung cancer cell line A549 are all commonly used in tumour cell studies and *in vitro* models of AGR2, all demonstrate expression levels >5-fold greater than that of the experimentally engineered cells. This is an essential characteristic of these model cell lines such that, it is now obvious that AGR2 protein is not being expressed significantly above the physiological level that we might expect from a transient transfection methodology. Rather, the relatively low level expression of AGR2 allows the interrogation of whether subtle expression of AGR2 reprogrammes cell signalling and/or signal transduction pathways.

Subsequently, the role of AGR2 following ER stress could then be assessed. AGR2 as an ER-resident PDI has previously been discussed (1.3.4.1), and ER stress response is published to impinge upon the function of AGR2 (179). The report by Higa *et al*. suggests that RNAi silencing of AGR2 restricts the unfolded protein response (UPR) via a reduction in the expression of Bcl-2 suppressor protein CHOP (CCAAT/-enhancer-binding protein homologous protein). In the synthetically engineered cell lines (Figure 3-5C) an enhanced CHOP response is apparent as a result of AGR2 expression relative to the AGR2-negative cells. Accordingly, we suggest that these cells are mimicking authentic AGR2-expressing cells in respect to ER stress responsiveness, and that the single gene introduction of AGR2 is enough to recapitulate the *in vivo* ER chaperone function of AGR2.

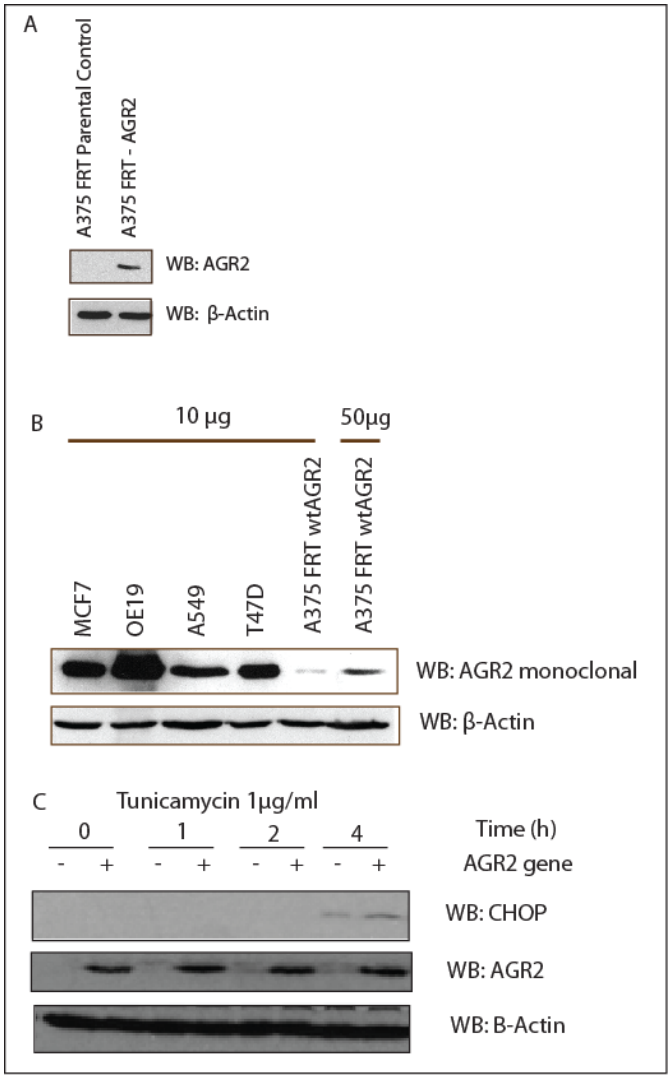


Figure 3-5 Characterisation of the A375-FRT-AGR2 experimental cell line examining expression of AGR2 and the effects of tunicamycin treatment on CHOP levels. (A) Immunoblot comparing the expression of AGR2 of the A375-FRT-AGR2 cell line, following engineering of stable expression in Figure 3-4, relative to the parental A375-FRT cells. (B) Comparison of the expression levels of the AGR2 isogenic cell population with authentic endogenous AGR2 expressing cell lines commonly used in vitro studies, MCF7, OE19, A549 and T47D. 10 µg of protein lysate from each cell line are shown, with an additional lane loaded with 50 µg of A375-FRT-AGR2 lysate. β-actin is shown as a loading control. (C) Measuring the integrity of the FRT-AGR2 by examining the effects of tunicamycin on the expression levels of CHOP. Previous studies identified the first clear evidence of the role of AGR2 in response to ER stress (179). A time course of tunicamycin at 1 µg/mL demonstrates that AGR2 presence elevates the expression of CHOP in response to stress (comparing lanes 7 and 8), consistent with the published reduction in CHOP levels induced by tunicamycin when AGR2 is silenced. These data suggest that the artificial A375-FRT-AGR2 cell line faithfully recapitulates the ER stress response.

Having established that the FRT-FlpIn cell system allows authentic recapitulation of the p53 response to genotoxic stress and AGR2 ER stress responsiveness, effects of AGR2 expression on cell growth were next investigated. The p53-suppressive nature of AGR2 (150) and effects of AGR2 overexpression on cell growth are broadly well documented (164;244), along with the influence of AGR2 on colony formation (175). It was therefore necessary to elucidate whether our isogenic, AGR2 over expressing cells could influence cell proliferation and invasion. A scratch assay (Figure 3-6) was applied (protocol devised and carried out by Dr. R. Hrstka, Masaryk Memorial Cancer Institute, Brno), in which a 100% confluent monolayer of the control and AGR2-expressing cells are scratched, and at regular intervals analysed to quantify the rate of wound healing (303). The data suggest that even subtle expression of AGR2 increases the rate of proliferation and migration compared to the AGR2-null cell line. Over the course of the 4 and 8 hour time points, AGR2-expressing cell migrate and proliferate into the wound significantly more readily than the non-expressing cell line, such that the wound is completely healed by 8 hours of growth of the AGR2-cells and this is not evident until 4 hours later in the parental cells.

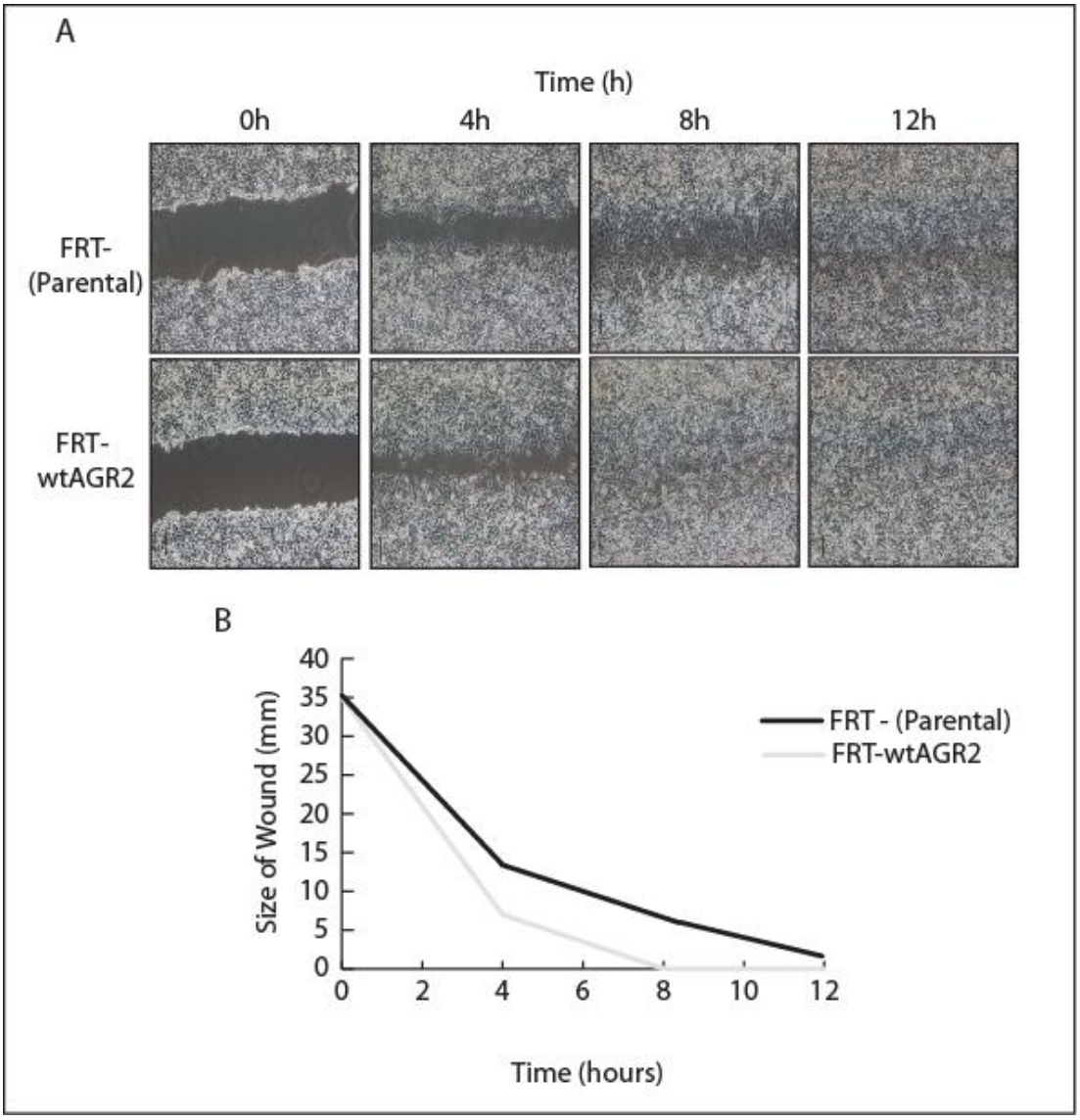


Figure 3-6 Analysis of A375-FRT-AGR2 cell proliferation and migration compared to the AGR2-null cell line, using a wound healing assay. A confluent monolayer of each cell line was scratched with a 10 μL pipette tip at zero time point. Images were then recorded at four hour time points for the duration of the study. Recorded images were then analysed by TScratch software (ETH) (304) to quantify proliferation.

3.2.5 The development of a panel of engineered cell models to define the extent to which AGR2 can alter proteostasis.

Having characterised the experimental engineered cell line in terms of p53 response, AGR2 responsiveness to ER stress, and cell proliferation/migration, we next set out to determine how a key domain of AGR2 can impact on cell signalling (Figure 3-7 and 3-8). As the only known domain of AGR2 that regulates localisation is the C-terminal ER-retention motif (187;188), we proceeded to engineer a panel of cell lines that either integrated the C-terminal ER-retention optimised AGR2-KDEL mutant (187;188), or the C-terminal deletion mutant, AGR2-ΔC, containing a four amino acid deletion of the wild-type ER retention motif (188) instead of the wild type non-optimal KTEL motif. Quantitative real-time PCR (qPCR) was employed in order to validate the presence (or absence, in the case of the control cell line) of the AGR2 isotype (Figure 3-7). These data demonstrate that in all of the experimental cases, the AGR2 gene is present (Figure 3-7A) compared to the FRT- parental cell (black line). Using the $^{\Delta\Delta}$Ct method of normalisation to GAPDH gene (Figure 3-7B), we can see that relatively similar levels of AGR2 and mutant mRNA are present in the engineered cells (Figure 3-7C), and that the AGR2 gene is absent from A375-FRT parental cells thus no mRNA is detected. Due to the isogenic nature of the cell lines, and the fact that all AGR2 constructs were recombined into the same A375-FRT clone, variation in mRNA levels may be as a result of a variety of factors including RNA degradation during extraction, mRNA decay, and hybridsation effects during PCR cycles.

Additionally, the C-terminal mutant panels were assayed for protein expression and localisation (Figure 3-8). Using confocal immunofluorescence microscopy to describe subcellular protein distribution, it was possible to identify populations of wild-type AGR2 protein (Figure 3-8A, top panel, [i]) frequently co-localised with the endoplasmic reticulum marker (protein disulphide isomerase, PDI) in a perinuclear location (Figure 3-8B, top panel). However, wt-AGR2 does not appear to always co-localise with PDI, characterised by some regions of red (488 nm) staining but lacking the green (594 nm) staining of the ER marker. AGR2-ΔC (Figure 3-8A, [ii] and 3-8B, middle panel) is minimally detectable within the cell, possibly only indicating

newly synthesised protein, as it is expected that the majority of the protein will be secreted into the extracellular milieu (188). Conversely, AGR2-KDEL mutant (Figure 3-8A, [iii] and 3-8B), exhibits a more prominent intracellular localisation and shows significant, almost complete, co-localisation with PDI. This indicates, as expected, that mutating the wild-type KTEL to the canonical KDEL motif locks the protein into the ER compartment. Finally, to ensure that AGR2-ΔC was in fact being trafficked out of the cell, and not rapidly degraded by cellular machinery, whole cell lysate and concentrated conditioned tissue culture media from wt-AGR2 and AGR2ΔC expressing cells were separated on a 12% polyacrylamide gel and immunoblotted for AGR2 using short and long exposure times (Figure 3-8C). These data confirm that a small proportion of AGR2-ΔC is present intracellularly, while the majority of protein is trafficked to the extracellular environment. Wild-type AGR2 is clearly apparent within the cell, and only a very small amount evident in the conditioned media. The latter data are consistent with a previous report that AGR2-ΔC is secreted (188) and further suggests that this engineered A375-FRT cell model reconstitutes an authentic AGR2 trafficking response .

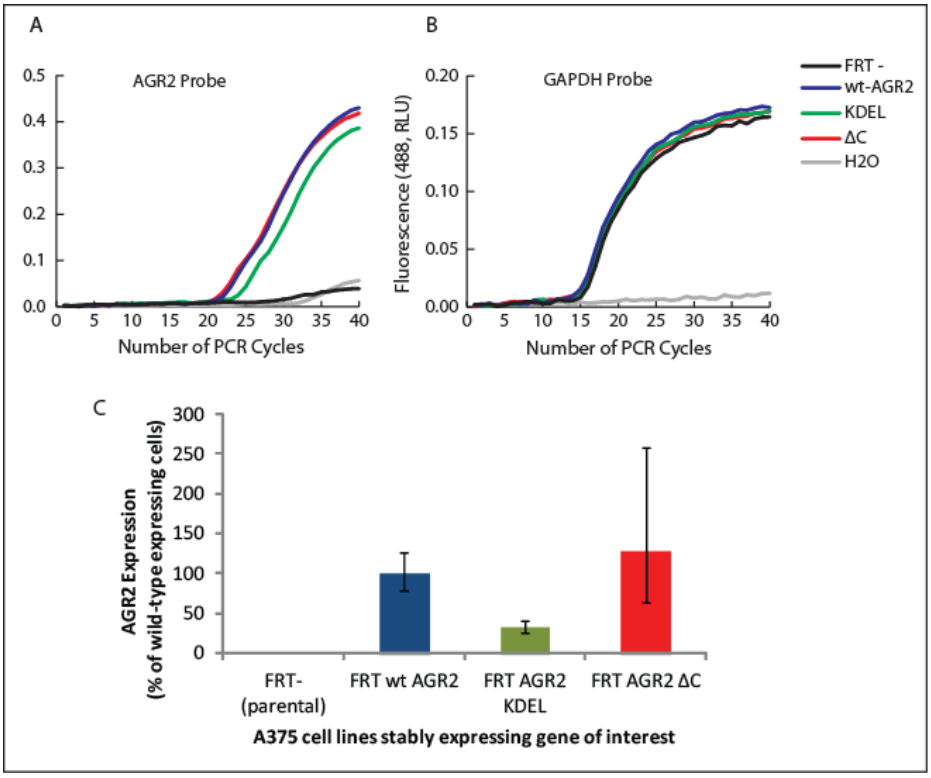


Figure 3-7 qPCR of C-terminal mutants of AGR2 compared to the AGR-null parental cell line. RNA was isolated from the four cell panels and the extent of (A) agr2 and (B) gapdh mRNA evaluated. The data are plotted as the hybridisation signal as a function of PCR cycle number and an average of three experiments. (C) Using the $^{\Delta\Delta}$CT method of qPCR data normalisation, an estimate of AGR2 mRNA for each cell line was estimated as a percentage of wtAGR2.

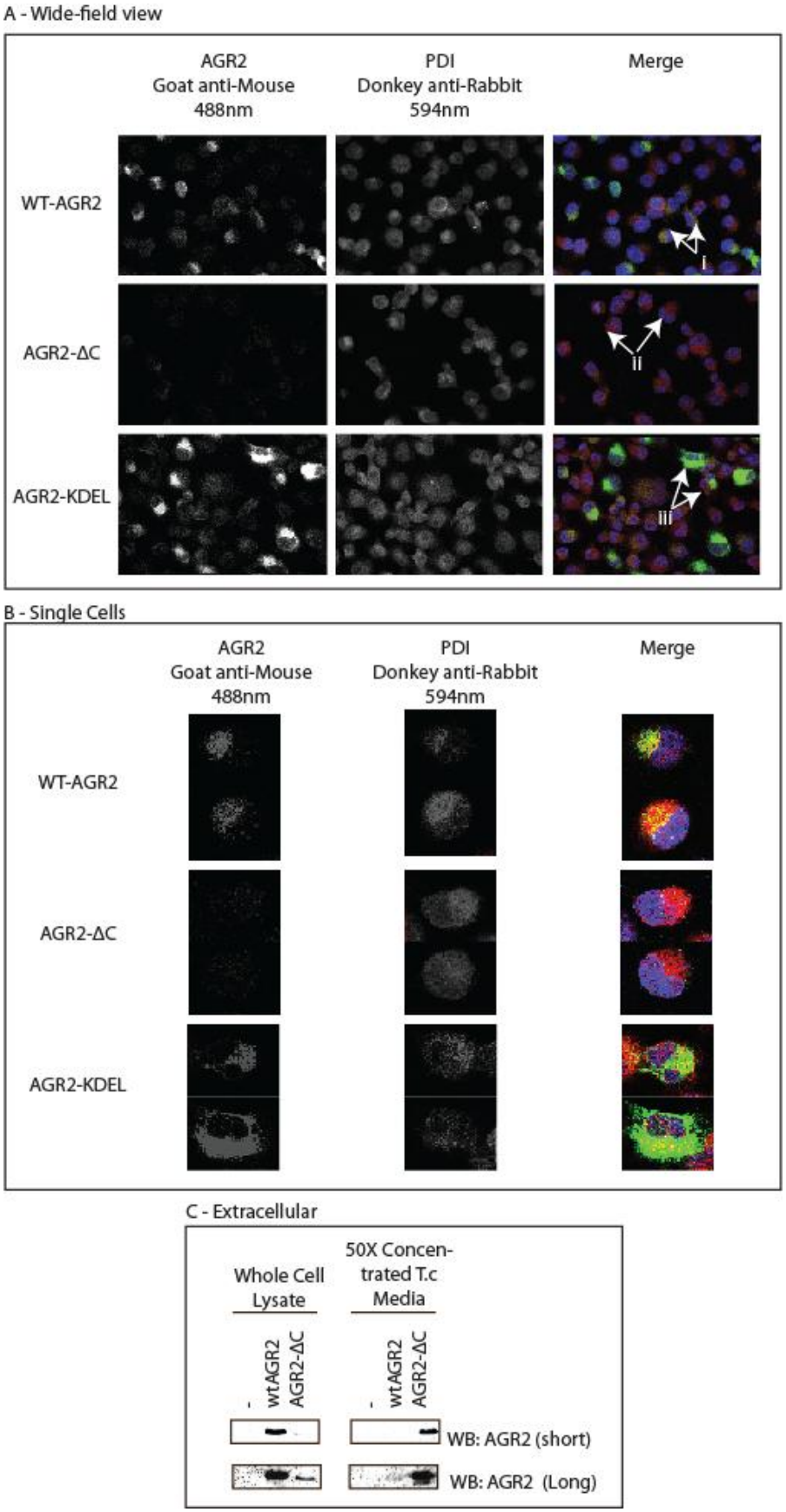

A - Wide-field view
AGR2
Goat anti-Mouse
488nm
PDI
Donkey anti-Rabbit
594nm
Merge
WT-AGR2
AGR2-ΔC
AGR2-KDEL
i
ii
iii
B - Single Cells
AGR2
Goat anti-Mouse
488nm
PDI
Donkey anti-Rabbit
594nm
Merge
WT-AGR2
AGR2-ΔC
AGR2-KDEL
C - Extracellular
Whole Cell Lysate
50X Concentrated T.c Media
-
wtAGR2
AGR2-ΔC
WB: AGR2 (short)
WB: AGR2 (Long)

Figure 3-8 Localised subcellular distribution of wt-AGR2, AGR2-ΔC and AGR2-KDEL mutant proteins using confocal microscopy and conditioned concentrated tissue culture media. (A&B) Cell lines were grown to ~70-80% confluency on glass cover slips before fixing and permeabilisation as per Methods. AGR2 mouse monoclonal antibody was used to probe for AGR2 proteins, and detected by goat anti-mouse secondary antibody conjugated to a 488 nm fluorophore. A rabbit polyclonal antibody protein disulphide isomerase (PDI, Enzo) detected by a 594 nm conjugated donkey anti-rabbit secondary provided an endoplasmic reticulum marker. The nucleus was stained using TOPRO-3 (Life Technologies). Microscopy was carried out using an Olympus FV1000 confocal microscope with a 60 × oil immersion objective lens and LAS-AF software. (C) Immunoblot of parental FRT, wt-AGR2 and AGR2-ΔC expressing cells demonstrate that, in lanes 1-3, wt-AGR2 is more easily detectable than AGR2-ΔC, which is minimally detected, within the cell lysate. By contrast, in lanes 4-6, when conditioned media from tissue culture was collected and concentrated 50-fold using 10 kDa molecular weight cut off filters, and loaded to a 12% polyacrylamide gel followed by AGR2 detection using an AGR2 monoclonal antibody the majority of AGR2-ΔC was evident secreted into extracellular milieu, compared to wt-AGR2 were only a very small amount was detected extracellularly.

Further characterisation of the wt-AGR2 and C-terminal mutant expressing cell lines employed the use of a novel cell growth assay to measure real-time changes in proliferation and cell doubling rate of engineered cells compared to the AGR2-null cell line (Figure 3-8A). The xCELLigence system (Roche), allows real-time label-free monitoring of cell growth by non-invasively measuring electrical impedance of cells cultured on gold micro-electrodes (305). As cell number increases, the resistance of the circuit correlates, resulting in a calculated quantitation of proliferation as cell index. Cells are seeded into microtiter plates, incorporating gold micro-electrodes, in equal number and in quadruplicate, over a period up to 96 hours (or until confluency has been achieved). Figure 3-8A[i] presents raw cell growth data of wild-type AGR2, AGR2- ΔC and AGR2-KDEL indicating that regardless of the subcellular localisation of AGR2, a cell growth benefit is achieved just through expression of this gene. Comparing wt-AGR2, AGR2-ΔC and AGR2-KDEL expressing cells with the AGR2-null cell line, all have an increased rate of cell proliferation [ii] and a reduced cell doubling time [iii]. For clarification that the process of stable cell engineering does not select for cells with increased proliferation rate, a further cell model is described (Figure 3-8B). The FRT-STOP cell line in this case expresses the N-terminal residues of AGR2, before truncating the gene product after only four amino acid residues (MEKI-STOP). This allowed an alternative control cell line which exhibits the integration of a gene, and selection agent resistance, without the gene product effect of gene expression. In this case, the FRT-STOP cell line did not provide any proliferative benefit to the cell line ([ii] and [iii]) compared to the wt-AGR2 expressing cells, and mimicked cell growth characteristics of the parental cell line. Collectively, these data indicate that the specific introduction of the AGR2 gene, irrespective of the subsequent subcellular localisation, promotes a proliferative nature of the cell line characteristic of an oncogene. The data observed are in agreement with previous studies highlighting the detection of AGR2 in multiple subcellular compartments in tumours, as regardless of location AGR2 appears to be promoting an increase in cell growth. However, it also conflicts with previous data suggesting that the C-terminal motif is essential for AGR2 function (188).

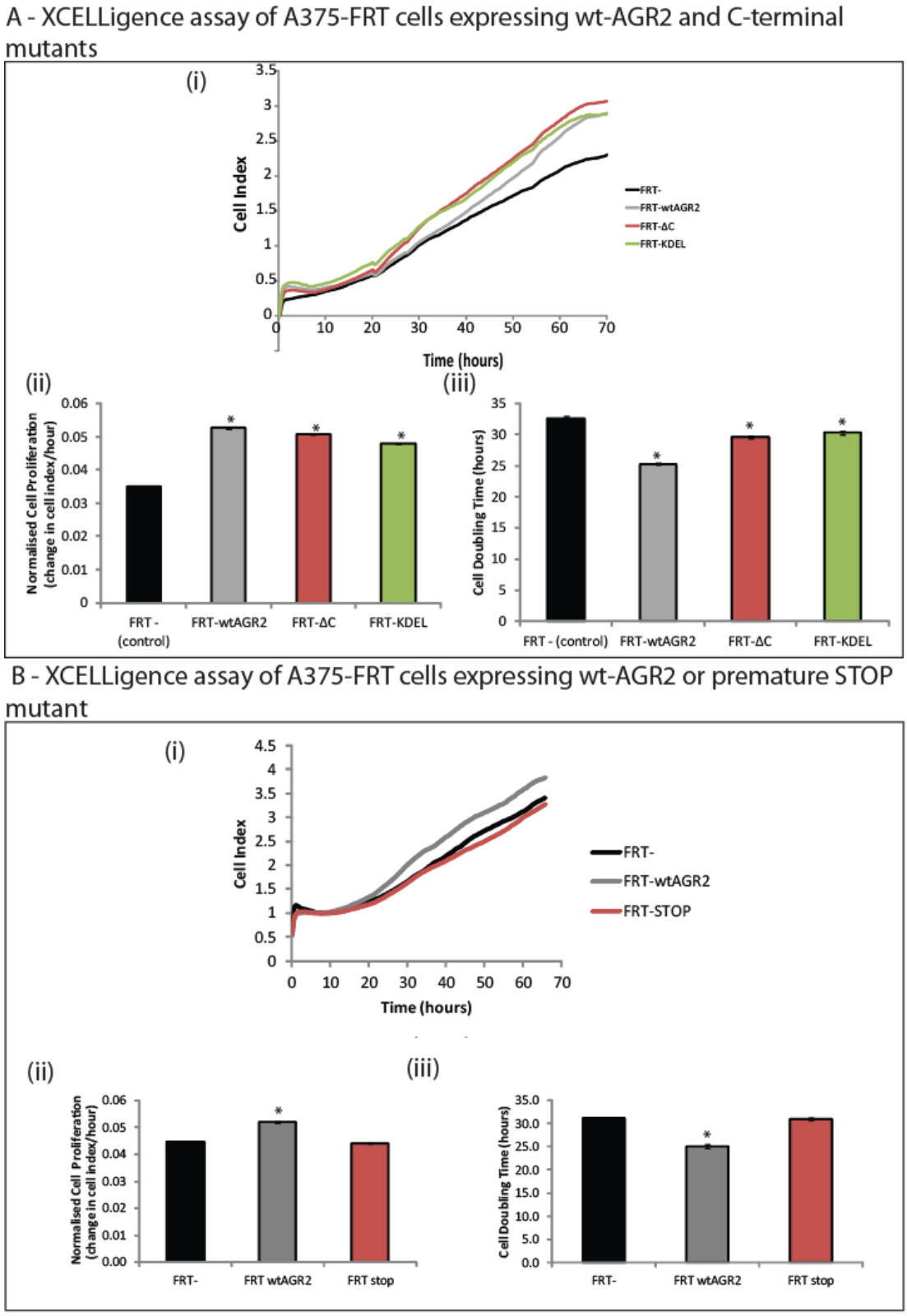


Figure 3-9 Analysis of the expression of the AGR2 protein, and its subcellular localisation, on cell growth using xCELLigence cell proliferation assay. 5000 cells were seeded into wells of a

microtiter plate with gold micro-electrodes (E-plate-VIEW, Roche) in quadruplicate and incubated at normal cell growth conditions (37^0C, 10% CO2) for up to 96 hours, or 100% confluency. Electrical impedance was measured in real-time every 15 minutes, mean cell index values were calculated and plotted as a function of time. (A) (i) A375-FRT cell index was compared to isogenic cells incorporating the wt-AGR2, AGR2-ΔC and AGR2-KDEL gene. Data analysis allows the calculation of (ii) cell proliferation (change in cell index as a function of time) and (iii) cell doubling time (the amount of time taken for the cell index to double). Indicating that the expression of the AGR2 gene in all cases promotes an increase in cell growth, despite differences in subcellular distribution (as seen in Figure 3-7). The asterisk (*) indicates significance $p<0.05$ in a two-tailed unpaired t-test. The experiment was carried out in quadruplicate, and data is representative of twice replicated experiments. (B)(i) Comparably, a FRT-STOP mutant expressing cell line was engineered to ensure that the selection of stably integrating cell lines was not biasing the proliferation of cells with increased growth rate. FRT-STOP cells expressed a truncated wt-AGR2 motif but the premature integration of a stop codon after the first four amino acid codons. This additional control cell line exhibited proliferation rate (i) and cell doubling (ii) characteristics similar to the AGR2-null parental cell line, and distinct from the wt-AGR2 expressing cell line. The asterisk (*) indicates significance $p<0.05$ in a two-tailed unpaired t-test. The experiment was carried out in quadruplicate, and data is representative of twice replicated experiments.

3.2.6 The effect of AGR2 expression on basal transcription of the cell

In order to define the dominant pro-oncogenic pathways that are altered by AGR2 expression, we first evaluated whether wt-AGR2 altered the basal transcription of the cell. There is evidence, indeed, that forced expression of AGR2 by ectopic expression or RNAi silencing of the endogenous protein from AGR2-positive cells can regulate EGFR pathway activation (194). This was achieved using an Illumina HT-12 BeadChips transcriptome microarray to screen the mRNA expression of > 25,000 genes using 48,000 probes. As controls for this experiment, cells expressing AGR2-ΔC, to control for extracellular effects (while also envisaging the use of data for the effect of secreted AGR2 as a mitogen affecting transcriptional pathways), or AGR3, to control for specificity, were engineered concurrently, using the same methodology and parental cell stock, with cells expressing wt-AGR2 (Figure 3-10). From each cell line, RNA was extracted and biotin-labelled in three independent studies to control for user bias and RNA degradation prior to hybridisation. Bioinformatic analysis deconvoluted complex data using Rank Product analysis with a 5% false discovery rate (Dr. Andrew Sims, University of Edinburgh). All transcriptomics data is provided in Appendix 1. Due to the reasoning that cell lines are isogenic, except for the expression of the exogenous experimental gene, only small specific changes in gene expression were expected. Broadly, the gene expression data (Figure 3-10A) demonstrate that the synthesis of wt-AGR2 has a negligible effect on global transcriptional homeostasis, presenting a very similar signature to the A375-FRT-parent cell and the AGR2-ΔC expressing cell, with only 27 genes expression diverging from the non-AGR2 expressing cell line (Figure 3-10B). Comparatively, the AGR3-expressing cell exhibits an almost 10-fold increase in gene transcription (229 genes) modulated in response to AGR3 presence. Of the 20 genes whose expression is altered following wt-AGR2 expression, all indicate significant up regulation, and one gene cluster of closely related genes demonstrates significant over-representation, the GAGE genes (family members GAGE2B , 4 (shared with AGR2-ΔC expressing cells), 5, 6, 12B,12C, 12F, 12G,12I and 12J). These are defined as cancer/testis genes, not expressed in normal tissues outwith the

highly tissue-restricted expression exhibited exclusively in normal testis and in a large proportion of tumours with varying histological origins (306).

Gene expression analysis alone is not generally accepted as proof of transcriptional effect unless further validation is sought. Due to the many steps and variations in microarray processing errors can be propagated due to background fluorescence, uneven hybridisation, fluorophore bleaching by light, temperature variation, hybridisation time and dye leaking (307). Therefore, a qPCR methodology was undertaken to validate data prior to further study of the effect of wt-AGR2 expression on the target genes (Figure 3-11). Somewhat surprisingly, the small cluster of GAGE genes that exhibited significantly higher expression in the wt-AGR2 expressing cells compared to the parental and AGR2-ΔC cells could not be reproduced by qPCR (Figure 3-11).

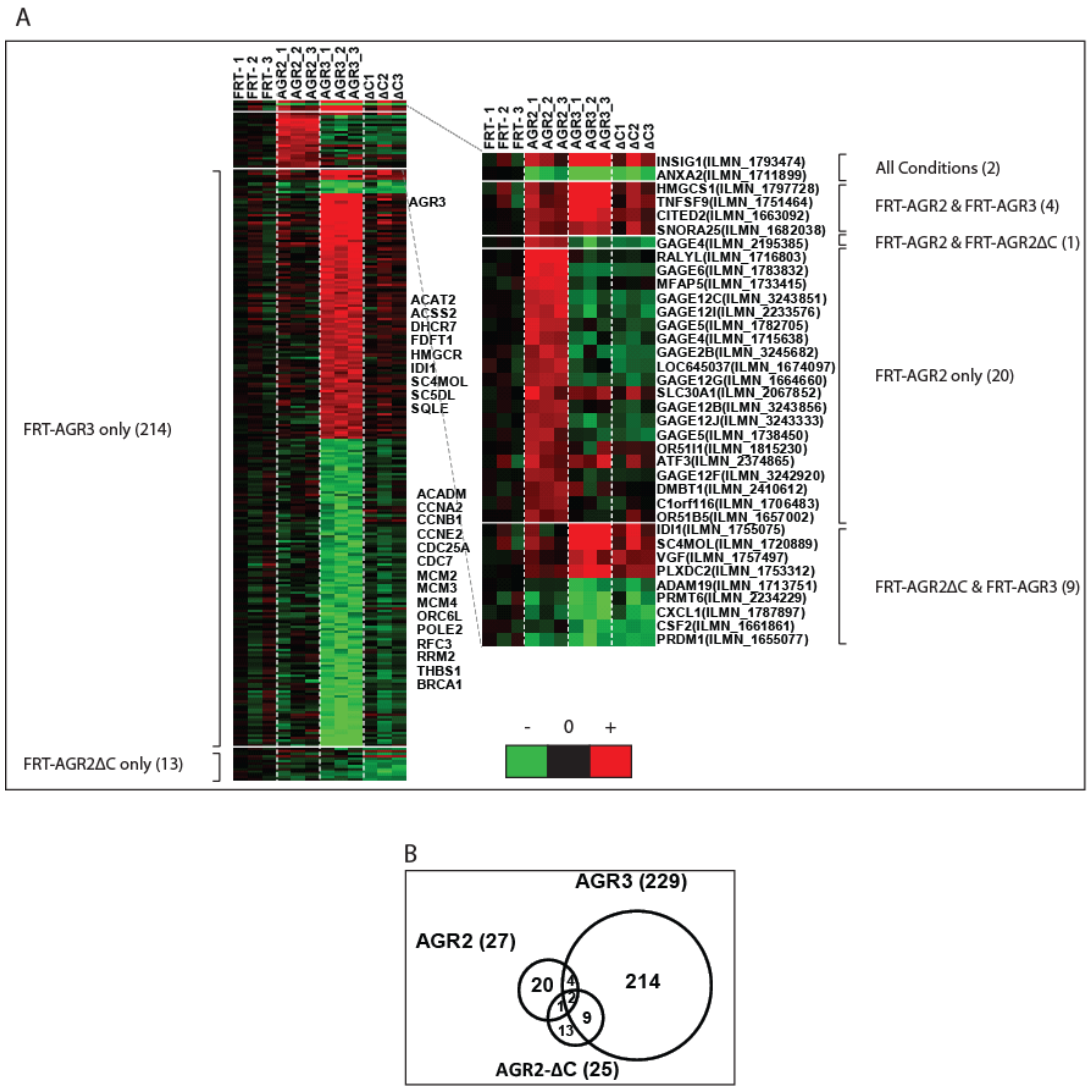


Figure 3-10 Analysis of transcriptional effects of AGR2 gene expression. (A) Heat map of differential expression of genes and gene probes from Illumina HT-12 v3 BeadChip microarray indicating biotin labelled isolated transcribed mRNA from the four conditions of parental A375-FRT cells, A375-FRT-wtAGR2, A375-FRT-AGR2ΔC and A375-FRT-AGR3 cells, reproduced in triplicate. Green indicates down regulation and red up regulation of hybridisation signal relative to the parental non-AGR expressing cell line mRNA. (B) Venn diagram demonstrating the overlap and individual gene expression changes, relative to the parental cell line, of isogenic cells expressing wt-AGR2, AGR2-ΔC and AGR3.

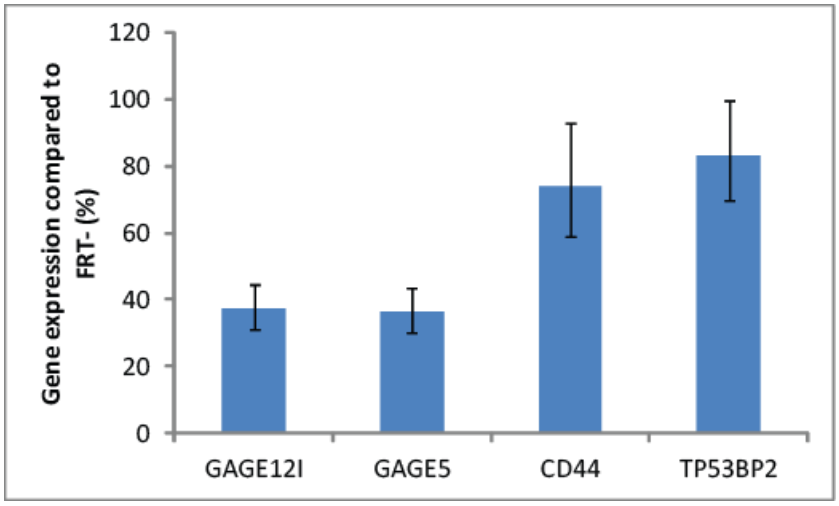


Figure 3-11 Normalised quantitative real-time PCR of selected genes identified from the Illumina HT-12 gene expression analysis (Figure 3-10), whose expression is up regulated following wt-AGR2 expression in isogenic cells relative to AGR2-null cells.

3.3 Discussion

The overexpression of AGR2 protein in a diverse range of tumour subtypes (205) makes it a compelling protein to begin to understand its role in cell growth, cell transformation and migration (164). Pro-oncogenic properties of AGR2 whereby it can exhibit these functions are thought to include its chaperone function as a protein disulphide isomerase in the ER, and the ability to mediate induction of oncogenic signals like epidermal growth factor (EGF) and the suppression of the tumour suppressor gene, p53 (150;188;194). However, mechanisms and co-factors employed to perform these roles are poorly understood. The majority of studies interrogating molecular basis of AGR2 signalling comes primarily from protein-protein interaction yeast-2-hybrid studies which have implicated C4.4A, α-DAG (172) and Reptin (199), and others requiring further validation (169). At present, interactomic studies have highlighted the complexity of AGR2 function, with roles in the ER (179), other subcellular compartments (187), membrane bound (204) and the extracellular environment (172). These studies, while informative and necessary, have not yet developed an understanding of the mechanisms of AGR2 protein overexpression in tumour cells and, mechanistically, how AGR2 influences the cell to exhibit pro-oncogenic characteristics.

Studies to date analysing the pro-oncogenic signalling of AGR2 firstly elucidated the p53-suppressive nature of AGR2 through an increase in specific phosphorylation on the tumour suppressor (150) and the nuclear exclusion of a key transcription factor involved in cellular homeostasis (187). The manner in which AGR2 performs these functions is not understood. Subsequent signalling pathway studies highlighted the induction of the EGF receptor ligand AREG, a growth promoting signalling protein, which could recapitulate the transformed phenotype absent when AGR2 was silenced (194). This study, by Dong *et al*., was the first to investigate and characterise the modulation of a tumour-associated signalling pathway by AGR2, implicating the nuclear localisation and dephosphorylation of the Hippo pathway co-activator, Yap-1, as the transcriptional activator to functionally link the activity of AGR2 to AREG

expression (194). However, the effect of AGR2 on AREG does not attribute for all the pro-oncogenic properties described by AGR2 overexpression. Therefore, the study presented here undertook a methodology to engineer biological tools which could ultimately be used in the characterisation of AGR2 signalling pathways, and more widely be used in the understanding of AGR2 in normal and diseased cells.

The engineering of a stable cell line, employing the Flp-In system, allows the creation of an isogenic cell panel, differing only in the single gene allele integration of the gene of interest. Initially, any cell line can be engineered to include the FRT site required for stable cell line generation, so it was considered which cell line provides the most appropriate model for analysis of AGR2 function. Although not essential, it was decided an AGR2-null cell line would be the most straightforward for downstream study, as any change in characteristic from the parental cell line could be attributed to the presence of AGR2. Theoretically, an AGR2 expressing cell line may propose a better model as the AGR2-dependent signalling pathways are already functional; however this makes analysis rather more complex such that external mechanisms would influence endogenous AGR2 expression levels resulting in potentially erroneous conclusions. It was also decided that the cell line should have a wild-type p53 background, due to the proposal of AGR2 as a p53-suppressor (150); this would give the greatest opportunity to study the role of AGR2 expression on one of the most aberrant homeostatic mechanisms prevalent in cancer studies. The A375 skin melanoma cell line met both these requirements, so provided the best model for AGR2 gene insertion study. There is a risk in the synthesis of any stable cell line panel, that the expression of the exogenous gene drives a selective pressure driving mutations in the cell signalling pathways. Such mutations could be propagated driving downstream detectable changes, which are then inaccurately attributed to AGR2 expression. To some extent these selective pressures cannot be completely avoided since the experiment requires the expression of detectable levels of the protein of interest. Therefore, in all circumstances where the A375-FRT cell model has been used in experimental protocols, passage numbers were restricted to a maximum of 10, and experiments comparing the expression of mutant or paralogous proteins were prepared from the same pool of A375-FRT parental cells.

Subsequently, control cells were always from this same pool also. The manner in which AGR2 drives selective pressure to promote potential oncogenic mutations over numerous cell doublings is a very interesting study in itself, however this was deemed outwith the scope of this study.

The gene recombination is integrated into the same location, and under control of the same promoter system, in all experimental cell lines synthesised resulting in a useful experimental model for the analysis of gene function. Alternative techniques rely on the transient transfection, and subsequent selection, of exogenous expression clones, or RNAi-based reagents to overexpress or silence the gene of interest. In these cases, the number of genes inserted into, or the extent of silencing, the host cell genome cannot be controlled. Therefore, these cell models may exhibit artefacts of over/under expression in protein folding, trafficking and ER stress. By using the recombination-based FRT genetic motif, it is known that only one gene can be controllably inserted into the host DNA, thus in further studies of mutants and paralogous proteins, divergences from the control are clearly as a result of the controlled variable. It must be noted that this technique is only of benefit if cells at each stage are appropriately characterised (as in Figures 3-1 and 3-2). Due to the random integration of the pFRT/*lac*Zeo gene into the host cell genome, some chromosomal location dependent effects on gene transcription and protein expression would be expected. Therefore, it is essential to correctly identify that only one FRT site is integrated into the genome (otherwise resulting in excision of intervening DNA if FRT sites are directly repeated in the same DNA strand, DNA inversion if the sites are in opposing orientations and deletion of genomic sequences) and that this integration does not have a dysfunctional effect on genes and signalling pathways essential to the study of the gene of interest (in this case, the p53-response to DNA damage, and the TLR3 interferon signalling system).

Following the recombination of the gene of interest into the host cell line (Figure 3-4), cells were assayed for described AGR2 characteristics. First of all, experimental cells express the AGR2 at the protein level in contrast with the parental cell line. This is an important outcome based on several factors: (i) The expression of AGR2 is not a lethal effect. As with any gene which disrupts normal cellular homeostasis, in

particular cell growth, the mere expression of the introduced gene may perturb cellular dynamics such that mechanisms within the cell sense erroneous cell signalling and action a senescent or apoptotic response. (ii) The gene expressed is translated into the gene product. It is plausible to consider that transcription of an exogenous gene, not normally exhibited by this cell line, may be subject to degradative machinery prior to phenotypic expression. This may occur at the mRNA level via salvage pathways, or during translation by proteosomal or lysosomal nutrient recycling systems. Fortunately, these engineered cells lack any apparent unstable characteristics, and exhibited the protein product at a level approximately $1/5^{th}$ of the expression level identified in commonly used cancer cell lines (Figure 3-5). Significantly, wt-AGR2 expressing cells also demonstrated an increased response to ER stress induced by a tunicamycin treatment time course, through an induction of CHOP mechanism (Figure 3-5). This supports the recently published RNAi silencing of AGR2 *in vitro* causing an increase in sensitivity to ER (179), suggesting that the artificially engineered FRT-wt-AGR2 cell line is faithfully recapitulating the ER-stress response observed in an endogenous AGR2 expressing HeLa cell line.

Further, the growth/invasion enhancement observed in the wound healing assay reiterates previous studies indicating that AGR2 over expression provides a survival benefit *in vitro* (150;175) and *in vivo* (218;231). There are apparent limitations to the wound healing assay, such that it is manual and subjective, and in the format used does not distinguish between cell proliferation and migration. Yet it provides a straightforward, rapid and economical technique (303), and coupled to TScratch analysis software (304), provides first line estimation of AGR2 expression effect on cell growth. These data were subsequently validated using a secondary xCELLigence real-time growth assay (Figure 3-9A). The benefit of cloning C-terminal mutant constructs of the AGR2 gene (Figure 3-7) was two-fold. It provided evidence to suggest that the A375-FRT cells were transporting the expressed protein to the correct subcellular compartment as suggested by bioinformatic analysis and by previous publications (188), detectable by immunofluorescence studies (Figure 3-8). They also provided excellent control cell lines for downstream analysis of wt-AGR2

specific activity, with the ultimate aim of determining whether the function was restricted to the ER-localised, the extracellular pool, or specific to the multi-compartment wild-type protein. While the mutant genes also being transcribed at a relatively similar level to that of the wild-type protein (Figure 3-7). Collectively, these data suggest that the isogenic cell panel described herein provides an appropriate model of AGR2 function in agreement with a range of publications of proposed AGR2 function.

The C-terminal mutant AGR2 expressing cell lines also provided an insight into the growth promoting nature of AGR2 using the xCELLigence cell growth assay (Figure 3-9A). A previous publication has described the wild-type C-terminal motif of KTEL being an essential requirement for the EGFR-activating activity of AGR2 (188;194). The data from these cell models somewhat conflict this previous study indicating that subcellular location is not a restraint on the proliferative growth effect of AGR2 expression since all AGR2 expressing cell models exhibited some level of increased cell growth rate, and reduction in cell doubling time. This raised the question as to whether the recombination of any gene was resulting in perturbation of normal cell growth. Therefore, a further cell line, incorporating a premature stop codon in the AGR2 gene, resulting in the expression a four amino acid residue peptide product, was engineered to challenge this (Figure 3-9B). As a result, since the A375-FRT-STOP cell line exhibited similar characteristics of the AGR2-null parental cell line, we can confidently conclude that the expression of AGR2 is the determining factor increasing the growth rate of these isogenic cell panels. The continuous cell monitoring nature of the xCELLigence proliferation assay makes it a more reliable measure of cell growth than the scratch assay (with particular reference to replicate experiments carried out in quadruplicate and statistical analysis of variation), yet this is balanced by increased cost, and technical skill required to prepare and analyse the assay.

The final characterisation of the wt-AGR2 expressing cell line was to analyse the effect of wt-AGR2 on the global basal gene expression of the cell (Figure 3-10). In order to appropriately determine changes in the transcriptome of AGR2 expressing cells, two additional controls alongside the AGR2-null cell line were assayed. The

AGR2-ΔC mutant as a control allows the identification of the reasoning behind the divergence of the C-terminal motif from KDEL (of ERP18) to the KTEL motif exhibited by wt-AGR2 which demonstrates reduced affinity for KDEL receptor proteins (192), and the properties shared between wt-AGR2 and secreted AGR2-ΔC. A375-FRT cells expressing the protein AGR3, with high levels of similarity, but less regularly identified as an oncoprotein in proteomic screens, is the final control as a control for specificity of AGR2 effects. Additionally, data collected from this cell line can be retrospectively studied to analyse the potential role of AGR3 whose overexpression, and uncoupling form AGR2, has been implicated in several subtypes of ovarian cancer (248), among others. Somewhat surprisingly, from a screen of 40,000 gene probes of the Illumina HT-12 BeadChip, only 27 genes were identified as exhibiting significantly different gene expression as a result of wt-AGR2 expression from those of the AGR2-null cell. Of these all were identified as having upregulated expression and bioinformatic analysis identified the GAGE gene cluster as being over represented. The GAGE gene family are a cluster of highly similar genes of repeated units which absent from normal tissue, except those of the testis, and highly expressed in cancerous tissue (308). This cancer specific nature of the GAGE genes has made them a highly attractive target for cancer therapy (309). The biological functions of GAGE genes are not yet clear, but emerging evidence suggests that they direct proliferation, differentiation and survival of human germs cells, and this may be recapitulated in cancer cells (310). Selected members of this group of genes, and others highlighted from the microarray, whose expression was modulated by AGR2 expression were assayed by qPCR as a validation, yet these could not be reproduced (Figure 3-11).

The inability of these gene expression experiments to reproduce the AGR2-dependent induction of the EGFR pathway, previously published (194), may be due to the relatively low level protein expression in the engineered cells relative to the authentic AGR2 expression in cancer cell lines (Figure 3.5B). Other tumour associated pathways AGR2 has been linked with perturbing include *c-myc*, cyclin D1 and survivin (231). This study, by Vanderlaag *et al*., found that protein levels of these three proteins, involved in cell survival, had reduced protein expression in

response to AGR2 silencing. Data presented herein indicate that overexpression of AGR2 in this A375 cell model does not affect transcriptional levels of these genes to a level of significance, and this can be justified by a number of reasonings: because AGR2 is not an endogenous protein in this system the signalling pathway to these proteins is not active or that AGR2 may indeed be affect the expression at the protein level as part of the function of an ER chaperone protein. Therefore, if threshold of AGR2 levels are important, then a hierarchy of AGR2 function would first impact on CHOP-dependent ER-stress response processes, prior to transcriptional-dependent effects on the EGFR pathway and other cancer associated signalling systems.

The engineering of the methodology and characterisation of these isogenic cell lines constitutively expressing AGR2, C-terminal mutants of AGR2 with distinct subcellular distribution and a closely related protein, AGR3, provides a novel, reliable and reproducible tool for the study of Anterior Gradient proteins. The exhibition of described AGR2-dependent functions in the suppression of CHOP and increased cell growth validate these cells as exhibiting displaying key natures attributed to AGR2. These tools therefore lend themselves to further interrogation, with particular respect to disease causing manifestations, of the functional roles of AGR2 *in vivo* with significance away from effects at the mRNA level.

Chapter 4: The Application of Quantitative Proteomic Methods to Identify AGR2 Reprogramming of Cancer Cells

4.1 Introduction

Since the cataloguing of the DNA blueprint comprising the complement of genes in the human genome, we have begun to understand that this knowledge base is conceivably immeasurably complicated throughout the central dogma of molecular biology, to reach the phenotypic level (Figure 4-1A) (311). That is the translated proteins that act as the functional effectors of the signalling pathways in response to changes in the cells' external environment. It was previously thought that these signalling pathways could be interrogated at the mRNA level, as signalling responses were thought to transcriptionally up- or down- regulate effector proteins. It has subsequently been found that this is not as simple as originally assumed, such that transcribed RNA levels do not necessarily correlate with protein expression levels (312;313); transcribed mRNA is not always translated into protein, and that the amount of protein transcribed from a specific amount of mRNA into protein can vary (314;315). The term 'proteome' was coined to describe the entire complement of proteins, incorporating modifications made, produced by an organism or system (316). Proteomics, in contrast to genomics, is dynamic and complex, in response to a changing environment, and adaptation to these changes (317). Proteins also lack the molecular capacity to be amplified, like DNA can be, as such no technology exists yet which can detect all proteins in the cell, so sequencing and identifying proteins requires a more challenging approach.

As such, significant investment has been made into developing technologies for interrogating the protein landscape to identify expression changes in diseased cells representing potential novel biomarkers (318). Undoubtedly, the impetus for progress in proteomics has been driven through revolutionary development in mass

spectrometry (MS) instruments, 'soft' ionisation methods and bioinformatical software packages (319). However, only a fraction of the proteins in the cell can be quantified by mass spectrometry. The detection of a peptide by MS is not reliant on abundance alone; rather the ionisation, chromatographic behaviour, solubility and interaction of the peptide ion with other compounds can all play a critical role. In the current preferred method, protein samples are first enzymatically cleaved into peptides, prior to fractionation of the sample by a variety of means (gel-based SDS-PAGE, subcellular compartment enrichment and strong cation exchange) before ionisation by techniques such as electrospray (320) or MALDI (matrix-associated laser desorption ionisation) (321). Ionised samples are analysed in a primary MS step, and precursor, or parent, peaks are isolated (Figure 4-1B). This provides a mass/charge (m/z) ratio for the whole peptide; however, this is insufficient for peptide/protein identification as several peptides will exhibit the same m/z ratio. Therefore, peaks are selected and the peptides are subjected to fragmentation, by collision with an inert gas, to identify the product or daughter ions. This second MS step fragments the peptide via the peptide bond into y and b ions, allowing comparison with a database of predicted masses of peptide sequences and subsequent amino acid residue identification (322;323) (Figure 4-1C). Thus, MS can identify the mass/charge ratio of the ionised peptide, and in a second MS step (tandem MS, or MS/MS), the peptide can be sequenced and, through the sequencing of multiple peptides, the protein identified through database searches (Figure 4-1B) (319).

Advances in the use of MS to study proteins have driven the development of methodologies to quantify the expression levels of proteins in experimental or clinical conditions. There are a range of quantitation methods available for the analysis of protein expression which utilise different technologies to differentiate between conditions giving information on the relative or absolute protein expression levels. The technique utilised in this study was stable isotope labelling by amino acids in cell culture (SILAC) (324). This methodology, pioneered by the lab of Matthias Mann (Max Planck Institute of Biochemistry, Munich), requires experimental cell lines to be grown in specialised media containing non-radioactive stable isotope labelled amino acids for metabolic incorporation into newly

synthesised proteins (Figure 4-2). Mammalian cells cannot synthesise a number of amino acids, therefore these essential amino acids must be provided in the tissue culture media, in this study we used the essential amino acid lysine, and the conditionally essential arginine (287;325). These can be radio-labelled, with ^{13}C and ^{15}N arginine and ^{13}C and ^{15}N lysine (R10K8, 'heavy') or ^{13}C arginine and ^{2}H lysine (R6K4, 'medium') or unlabelled (R0K0, 'light') and supplemented into the media as free amino acids to support growth, and in turn are incorporated into nascent protein chains during translation. Controlled absence of any unlabelled specific amino acid is imperative, therefore cells must be cultured in media containing dialysed serum, and care must be taken to avoid enzymes regularly used in tissue culture, such as trypsin, which might provide a source of unlabelled amino acids. Following a number of cell doublings, each occurrence of the specified amino acid will be replaced by its isotopically tagged analogue. As there is no chemical difference between the natural and labelled amino acid (SILAC labelling has been tested in a range of cell lines and primary cultures with no detrimental effect on growth, morphology or biological activity (326)), the cell should act identically to a control cell line grown in media with natural amino acids.

Therefore, labelled cells can be experimentally treated along with an untreated control, subsequently harvested, and because the label is encoded directly into the amino acid sequence of every protein, the extracts can be mixed, and processed (reduced and alkylated) concurrently removing additional potential sources of error (324). Mixed lysates are then subject to enzymatic digestion, most commonly and in this study, with trypsin. Due to the specificity of the trypsin enzyme, cleaving only when the carbonyl group is contributed by a lysine or arginine residue, every peptide, excluding the very C-terminal peptide, will therefore incorporate an isotopic tag (327). In shotgun proteomics, the trypsinised peptides are then injected directly into a high performance liquid chromatography front-end, or alternatively to reduce sample complexity we fractionated the protein lysates off-line using gel electrophoresis (326) prior to sample introduction. This additional fractionation aims to allow maximal identification of proteins within the sample, and also aid the identification of lower abundance peptides as the fractions will be less complex

(328;329). Therefore, once tryptic peptides are injected to the MS, a mass shift between cells cultured in the distinct labelled amino acid containing conditions exists and can be detected in the first MS. Following the MS/MS peptide identification and confirmation, the area under the MS1 spectra can be integrated and a calculation of peptide quantitation derived. Using the quantitation of several unique peptides, a measure of protein expression level (and confidence in this expression) can be compared to the differentially labelled cell lines, resulting in relative quantitation of the protein. The early combination of samples and metabolic labelling, results in SILAC being likely the most accurate quantitative MS method, and is particularly suitable for assessing comparatively small changes in protein levels (330).

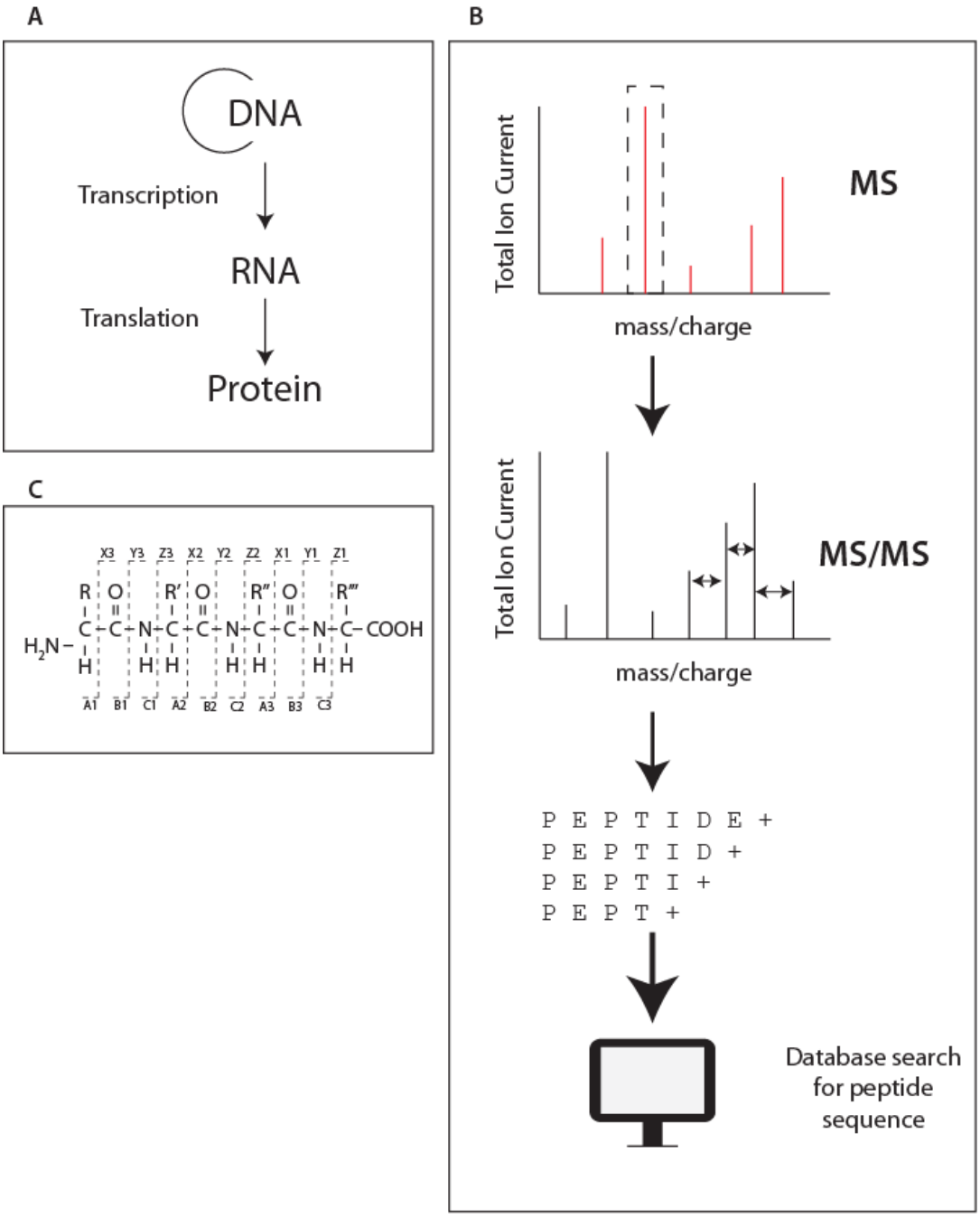


Figure 4-1 Overview of mass spectrometry based study of proteomics. (A) The central dogma of molecular biology where static, unchanging DNA in the genome is transcribed to mRNA, before translation into the final gene product, the protein. (B) Illustration of mass spectrometry application in peptide, and subsequently protein identification. (C) Fragmentation at the peptide bond, yielding different ion series

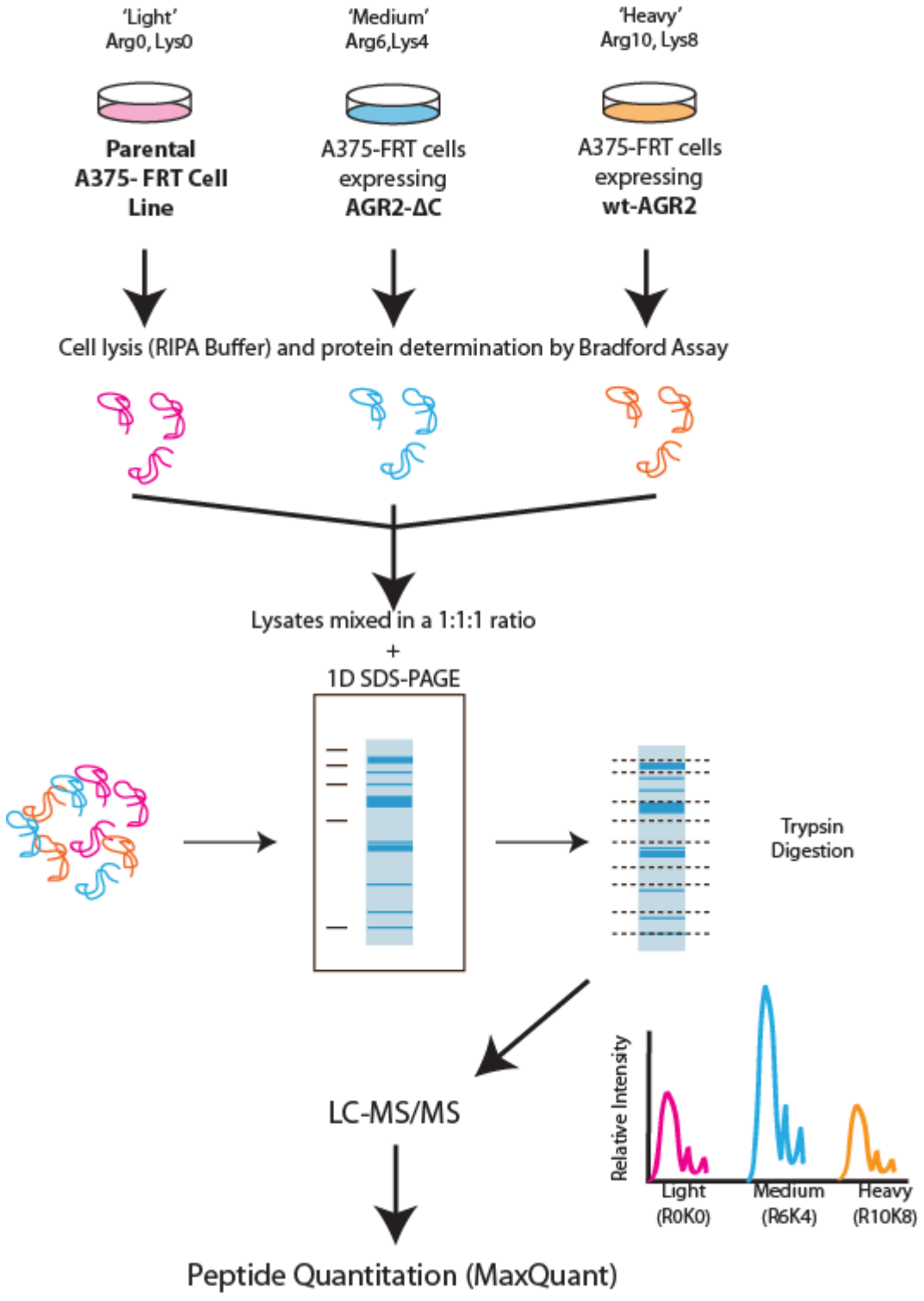


Figure 4-2 Stable isotope in amino acid cell culture (SILAC) protocol employed to interrogate how wtAGR2 expression affects the protein landscape.

Thus, SILAC can be appropriately utilised in the identification of expression changes induced by the expression of the gene of interest and can then be mined for over-representation indicating the specific action of the experimental gene using a range of widely available bioinformatics packages (331). The software package Ingenuity Pathway Analysis (IPA, Ingenuity Systems, www.ingenuity.com), allows the identification of biological mechanisms, pathways and functions most relevant to a high-throughput proteomic data set to allow interpretation of the biological significance and meaning of variation in expression levels. IPA incorporates a curated distinct knowledgebase collating literature of biological and chemical knowledge from published peer-reviewed journals and biomedical databases to allow computational disentangling of complex data to identify key modulated biomolecular events.

These techniques, coupled to the engineered isogenic stable cell lines, described and characterised in Chapter 3, allowed the interrogation of the cell signalling pathways disrupted or reprogrammed by the low level expression of the AGR2 gene. The coupling of novel quantitative mass spectrometry techniques with isogenic cells lines is a somewhat new application for SILAC in the understanding of protein function and signalling, with relatively few publications to date (332;333). As AGR2 is fundamentally an ER resident PDI (179), perhaps its chaperone function has a primary effect on cell signalling. As such, we can evaluate whether the addition of the single AGR2 expressing allele can reprogramme the cellular steady state proteome under conditions which the transcriptome changes were negligible (Figure 3-10).

4.2 Results

4.2.1 SILAC quantitation of the protein landscapes of isogenic cell lines incorporating the AGR2 gene

The SILAC methodology allows a degree of multiplexing of the experiment design, to incorporate relevant controls for the analysis of function of the AGR2 gene of interest. We employed the triple-SILAC technique (Figure 4-2), allowing the relative expression of three isogenic cell lines to be compared, to compare the A375-FRT (AGR2-null) parental cell, A375-FRT-wtAGR2 incorporating the wild-type AGR2 gene, and the A375-FRT-AGR2ΔC which recombined the mis-localised mutant AGR2 (Figure 3-8) protein as a control for specificity. Due to the cells being isogenic, any modification from the parental cell line suggests a potential role of the introduced gene in the reprogramming of the cell. Control AGR2-null cells were cultured in media containing unlabelled lysine and arginine residues ('light' condition), wild-type AGR2 expressing cells incorporated the 10 Da labelled arginine ('heavy' condition), and 8 Da lysine, while AGR2-ΔC expressing cells the 6 Da arginine and 4 Da lysine ('medium' condition) prior to gel-based separation and MS processing, thus peptides detected from each condition can be distinguished (Figure 4-2). Figure 4-3A presents an illustration of the distinct MS1 spectra expected of peptides derived from the three conditions and the mass/charge shift expected from lysine or arginine containing peptides. Raw mass spectral data (Figure 4-3B) presents an example of MS1 data of a peptide identified in two of the conditions, light and medium labelled, of a lysine containing peptide whose product ion is separated by a 4 Da m/z shift. From MS1 spectra, it can be identified that the peptide must contain a C-terminal lysine (due to the 4 Da shift) and that there is a small variation in the detection of the peptide with less peptide abundance in the medium labelled condition. The product ions of both conditions were isolated and fragmented, allowing sequencing of the peptide as DSYVGDEAQSK. Submission to bioinformatical databases allowed this peptide to be identified as unique to the cytoskeletal protein, actin (IPI00021440, ACTG1), and, together with a number of

other peptides, allowed statistical quantitation of the protein expression levels in the two conditions and a relative expression level derived.

MaxQuant quantitation (289) utilises the generated peak list, SILAC label and extracted ion current-based quantitation to calculate a posterior error probability (PEP), and false discovery rate based on the search engine (Mascot) results, peptide to protein group assembly allowing filtration and presentation of data. 29044 peptides were sequenced in the experiment allowing quantitation of 2877 proteins, and of these 2425 proteins were present in all 3 conditions, and thus used for relative quantitation (Appendix 2). Data were then imported to Microsoft Excel for further analysis. Due to the cell lines being isogenic, except for the recombination of the experimental gene, few significant changes from the expression of the control AGR2-null cell line were anticipated. Indeed the expression ratio of most proteins remains unchanged (Figure 4-4A) and by plotting natural logarithm ratios of protein expression changes when wt-AGR2 is expressed compared to the AGR2-null cell, against the $-\log_{10}$ PEP (Figure 4-4B), the majority of protein expression levels quantified across all three conditions were not significantly altered more or less than the empirically derived 20% change cut off, from the AGR2-null cell line. This suggests that AGR2 is not completely abrogating functional signalling pathways and that the pro-oncogenic, pro-growth reprogramming of the cell by AGR2 is much more specific. The expression of wildtype-AGR2 resulted in the change of expression of 735 proteins (333 up-regulated and 402 down-regulated) by >20% (Figure 4-4C). By comparing the perturbed expression of proteins when wt-AGR2 is expressed, to expression levels of proteins from the secreted AGR2-ΔC expressing cells (Figure 4-4C), we can distinguish that wt-AGR2 is acting specifically within the cell to perturb signalling pathways, as there is only a small proportion of overlap between the proteins whose expression changes as the gene is exhibited. Further, these overlapping data of the 226 proteins whose expression is linked to both wt-AGR2 and AGR2-ΔC (91 upregulated and 135 downregulated) may subsequently be useful in identifying the extracellular function of AGR2.

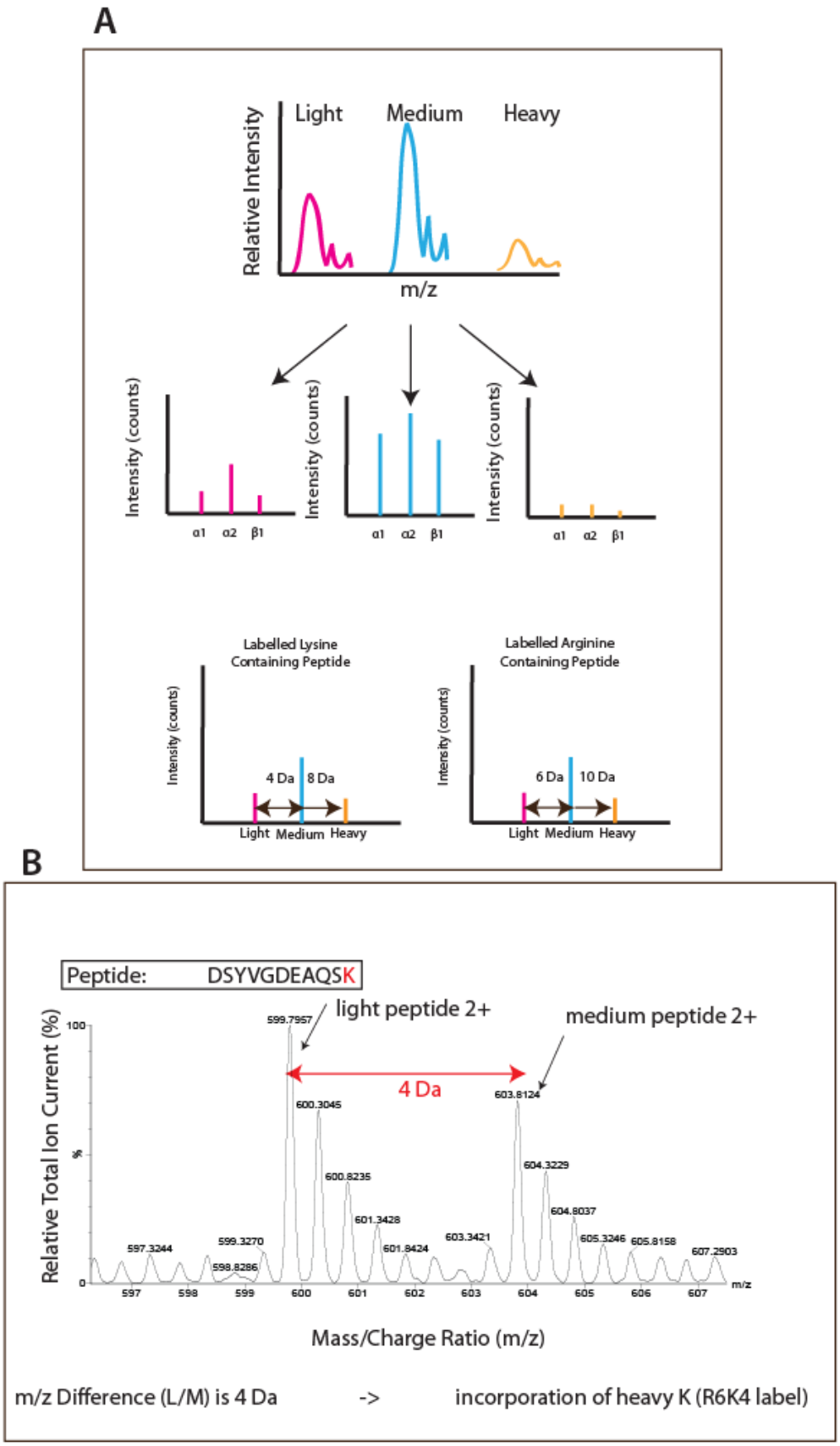

A
Light
Medium
Heavy
Relative Intensity
m/z
Intensity (counts)
α1 α2 β1
Labelled Lysine Containing Peptide
4 Da 8 Da
Light Medium Heavy
Labelled Arginine Containing Peptide
6 Da 10 Da
B
Peptide: DSYVGDEAQSK
light peptide 2+
medium peptide 2+
4 Da
Relative Total Ion Current (%)
Mass/Charge Ratio (m/z)
m/z Difference (L/M) is 4 Da -> incorporation of heavy K (R6K4 label)

Figure 4-3 Example of SILAC Quantitation. (A) Illustration of the stable isotope labels used in the labelling of conditions and the m/z shift expected from the incorporation of the distinct tags on lysine and arginine residues. (B) Simplified mass spectra of the peptide DSYVGDEAQSK from the protein Actin (IPI00021440, ACTG1) present in A375-FRT and A375-FRT-AGR2ΔC exhibiting a 4 Da m/z shift and variation in peptide detection.

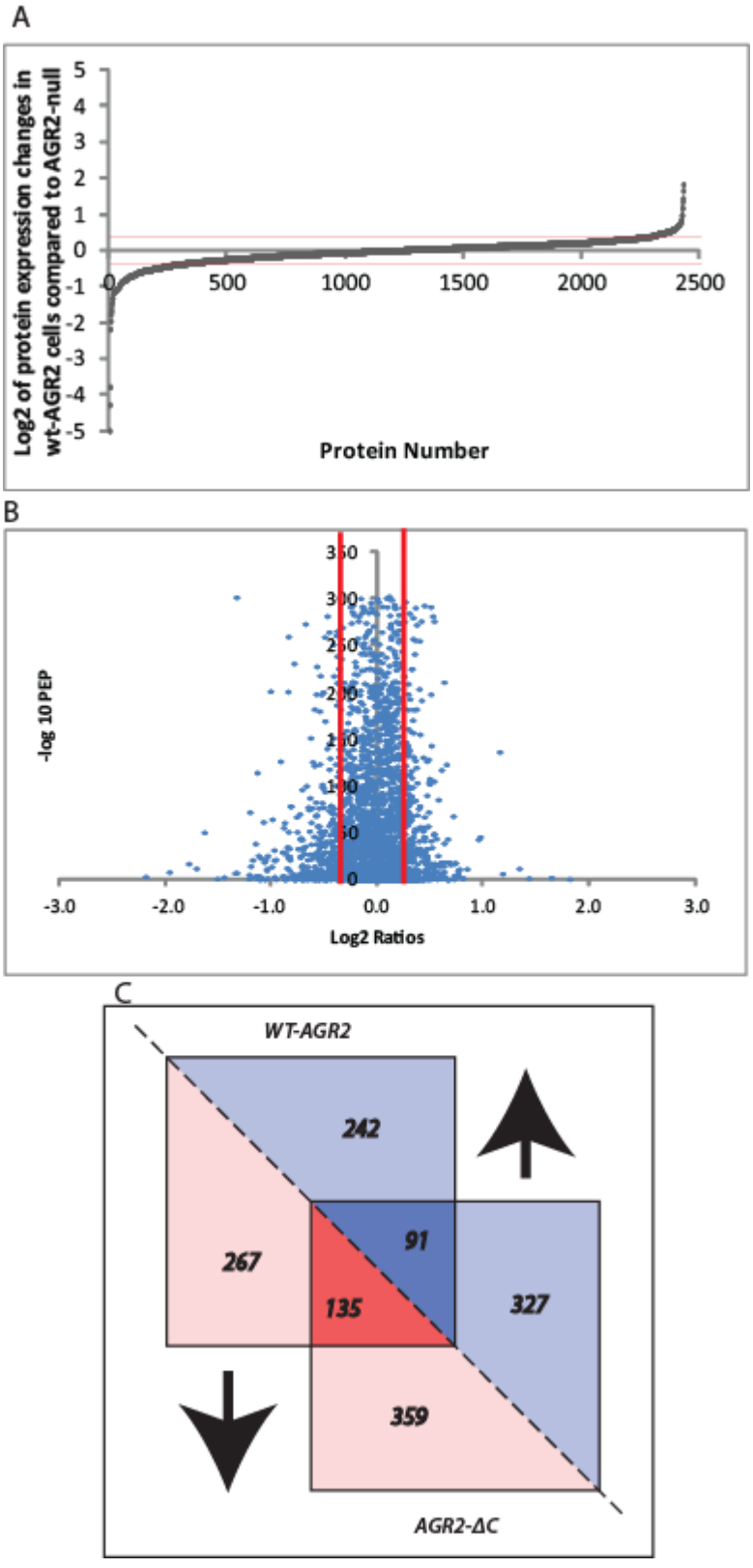


Figure 4-4 Quantitative mass spectral data indicates that wt-AGR2 expression subtly reprogrammes the proteome of the cell. (A) Log2 of the expression ratios of the 2425 SILAC

quantified proteins of wt-AGR2 expressing cells compared to isogenic AGR2-null cell line, demonstrating the majority of expression is clustered near the x-axis indicating no significant change. (Red lines indicated a 20% expression change gate). (B) Plot of Log2 ratios of protein expression changes induced by wt-AGR2 relative to AGR2-null cells plotted against log10 of the posterior error probability. The red lines indicate the threshold for change in expression set at ±20%. (C) Venn diagram indicating the number of proteins up- and down-regulated (>20%) as a result of wt-AGR2 or AGR2-ΔC gene recombination relative to the parental cell line lacking any AGR2 gene, and the overlap of shared expression profiles.

4.2.2 Functional analysis of SILAC proteomic data

As is the case with all OMICs technologies, the value of the data lies in the functional interpretation of the results to drive a phenotype or disease pathway. Also, proteomic data is inherently complex as each of the numerous detected proteins may be a member of several pathways, connected by multiple protein interactions and regulated by a variety of remote regulators. Knowledge based approaches deal with this complexity by creating a structured database of protein interactions, pathways and protein-disease associations from experimental literature and a set of statistical tools to compare the proteomic profiles with this rich source of accumulated knowledge (334). The knowledge database incorporated into Ingenuity Pathway Analysis avoids operator bias and blindly assumes no previous knowledge of the experiment, allowing visualisation of high throughput data in the context of biological networks. Submission of protein expression ratios for wt-AGR2-expressing cells versus AGR2-null parental cells which were quantified using at least three peptides (to increase reliability of data set (324)), and exhibited expression changes of at least 20% were submitted to IPA for evaluation (data processing carried out by Dr. Judith Nicholson, Australian Proteome Analysis Facility, Macquarie University, Sydney). This bioinformatic analysis identified TP53 transcription factor activity as the most perturbed (p-value = 1.42×10^{-4}), through the aberration of 39 different proteins whose expression is linked to p53 signalling, and that this transcriptional pathway was in a state of inhibition (Table 4-1 and Figure 4-5). TFEB (p-value = 1.16×10^{-3}) and E2F1 (p-value = 2.20×10^{-4}) pathways were also identified as being significantly inhibited and activated, respectively. This result links to the increased growth of wt-AGR2 expressing cells (Figure 3-6 and 3-9), yet the misregulation of these pathways were not previously highlighted by the whole genome gene expression profile (Figure 3-10). Comparatively, expression changes as a result of AGR2-ΔC gene insertion did not show any significant effect on p53, or significant modulation of any other transcriptional pathway (Table 4-2).

Table 4-1 Pathway annotation using Ingenuity Pathway Analysis highlighted the dominant pathway suppressed by wt-AGR2 expression was TP53 transcriptional regulator.

Transcription Regulator	Number of protein expression changes	p-value of overlap	Predicted Activation State	Regulation (fold)
HNF4A	93	1.70×10^{-5}		
TP53	39	1.42×10^{-4}	Inhibited	-2.747
E2F1	21	2.20×10^{-4}	Activated	2.204
E2F2	7	6.73×10^{-4}		
TFEB	5	1.16×10^{-3}	Inhibited	-2.201

Table 4-2 Transcriptional regulator pathways are not affected by the expression of AGR2-ΔC

Transcription Regulator	Number of protein expression changes	p-value of overlap	Predicted Activation State	Regulation (fold)
HNF4A	126	6.18×10^{-8}		
SREBF1	13	3.81×10^{-5}		-1.517
MCYN	21	6.90×10^{-5}		1.678
IRF2	6	1.04×10^{-3}		-0.044
MYC	19	1.71×10^{-3}		

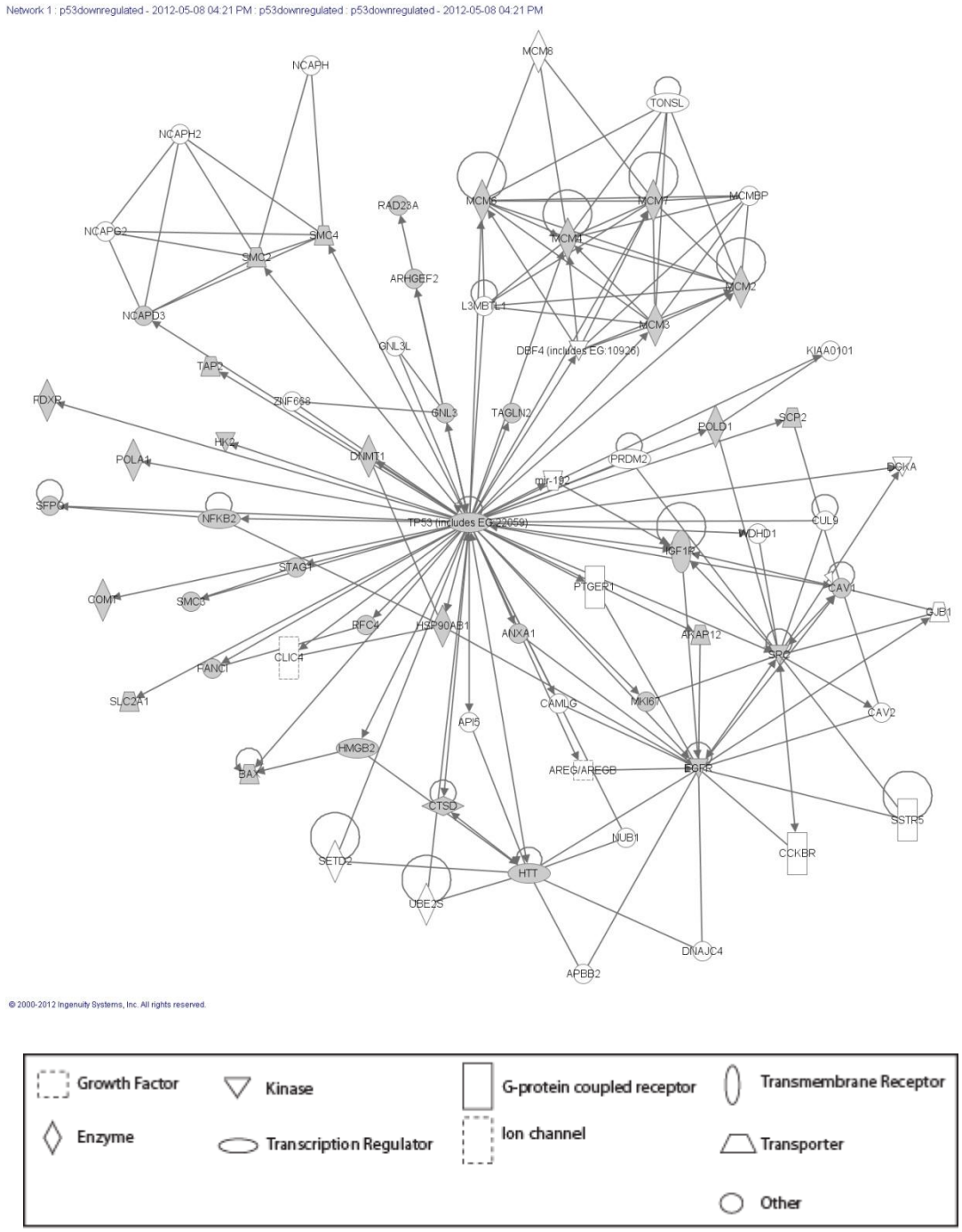


Figure 4-5 Ingenuity network map of TP53 signalling axis. Nodes highlighted with grey shading are significantly changed following the overexpression of wt-AGR2 in A375-FRT cells, and thus included in the submitted dataset

4.2.3 Validation of p53-dependent activity repression in wtAGR2-expressing A375-FRT cells

The bioinformatic elucidation of suppression of the p53-signalling axis in wt-AGR2 expressing cells was a stimulating result when compared to available literature demonstrating previous reporting of the p53-suppressive nature of AGR2 (150;187). However, due to the nature of *in silico* techniques and lack of appropriate replicates, biological validation of the hypothesis was sought for assurance. We had previously used the platinum containing, DNA-crosslinking cancer therapeutic cisplatin to demonstrate that the integration of the recombination target into the host cell line had not disrupted the functional DNA damage response pathways in line with previous publications (297) (Figure 3-2). This demonstrated that p53 was induced, following DNA damage, promoting cell repair or a pro-apoptotic response. AGR2 is also described as mediating cisplatin resistance in xenografts (248). Accordingly, the panel of AGR2-null isogenic cells, and cells expressing wild-type, KDEL or ΔC mutants of AGR2, were challenged with cisplatin over a 24 hour time course. Thus allowing for evaluation of whether or not AGR2 expression indeed had any effect on the p53-response to DNA damage, to recapitulate the SILAC-IPA coupled data analysis (Table 4-1). Additionally, the C-terminal mutant cells might suggest whether any deviation was as a result of the ER localisation of AGR2 (KDEL, untested) or extracellular secreted (ΔC, SILAC analysis had not highlighted p53-signalling to be perturbed, so the hypothesis expected this not to be the case).

Western-blot analysis of the p53-response to cisplatin-induced DNA damage in the isogenic cell panel, followed by densitometry measurements of protein bands highlighted, in the first instance, that wild-type AGR2 expression was indeed suppressing the activation of p53 over the time course (Figure 4-6A, upper panel lanes 1-7 compared to lanes 8-14). p53 activation was estimated to be approximately one quarter of that of the AGR2-null cell line after 24 hours treatment (p53/β-actin densitometry, 21.2 compared to 5.0 relative units). Moreover, not only were the expression levels of p53 suppressed in wt-AGR2 expressing cells, but p53-dependent activity read out of p21, was also significantly decreased compared to the AGR2-null cells. Comparatively, the C-terminal mutants of AGR2 did not exhibit the same p53-

suppressive nature identified in the wild-type expressing cells and p53 was induced in a manner similar to the AGR2-null cells (Figure 4-6A, lower panel compared to upper panel), suggesting that the wild-type C-terminal motif is essential to the p53-suppressive nature of AGR2. For complete analysis, an end point comparison of no treatment and 24 hour cisplatin treated lysates of the cell panel were prepared (Figure 4-6B). This established that at basal levels, untreated wt-AGR2 expressing cells did not significantly inhibit p53 expression levels detectably, however, p21 expression was estimated to be almost 50% reduced compared to the AGR2-null parental and AGR2 mutant-expressing cells (Figure 4-6B, lane 2 versus 1, 3 and 4). So despite not directly affecting basal p53 levels, endogenous p53-dependent activity was being attenuated. Following cisplatin treatment, wt-AGR2 expressing cells exhibited decreased p53 expression, p53 ubiquitination and subsequently p21 levels (Figure 4-6B lane 6 vs 5, 7 and 8) compared to AGR2-null and cells expressing mutant AGR2. These data suggest that only wt-AGR2 expressing cells can suppress p53 and p53-dependent activity, consistent with the Ingenuity Pathway Analysis of the triple SILAC screen.

Taking into consideration that the A375-FRT cell panel is a model used for the analysis of AGR2 over-expression and function, it is imperative to translate these conclusions into endogenous AGR2-expressing cancer cell lines to challenge whether the p53/p21-suppressive activity of AGR2 is an authentic signalling activity relevant to the cancer cell phenotype. To avoid inducing stress mechanisms related to protein over-expression by transient transfection; siRNA-induced gene silencing of AGR2 was employed to suppress the expression of endogenous AGR2 in two different cell lines, allowing evaluation and recording of any effect on p53/p21 signalling axis. Consistent with previous data described, siRNA-mediated depletion of endogenously expressed AGR2 protein in MCF7 or A549 cancer cell lines (both having a functional wild-type p53 pathway) resulted in steady state p53 protein elevation at time points 72 hours and 96 hours post-siRNA transfection (Figure 4-7). After 72 hours of AGR2-targeted and non-targeted nonsense siRNA treatment in both cell lines (Figure 4-7A and 4-7B lane 1 compared to 2), AGR2 levels were significantly suppressed. Coupled to this, p53-levels were induced in the AGR2

silenced cells, as was the level of p53-dependent p21 activation. Additionally, after 96 hours treatment, this trend was still observable (lane 3 vs 4).

Collectively, these data suggest that the p53-suppressive nature identified by the SILAC proteomics screen and *in silico* bioinformatical analysis could be validated by biochemical methods, and indeed held relevance when the hypothesis was tested in endogenously expressing AGR2 cancer cell lines exhibiting a wild-type p53 pathway.

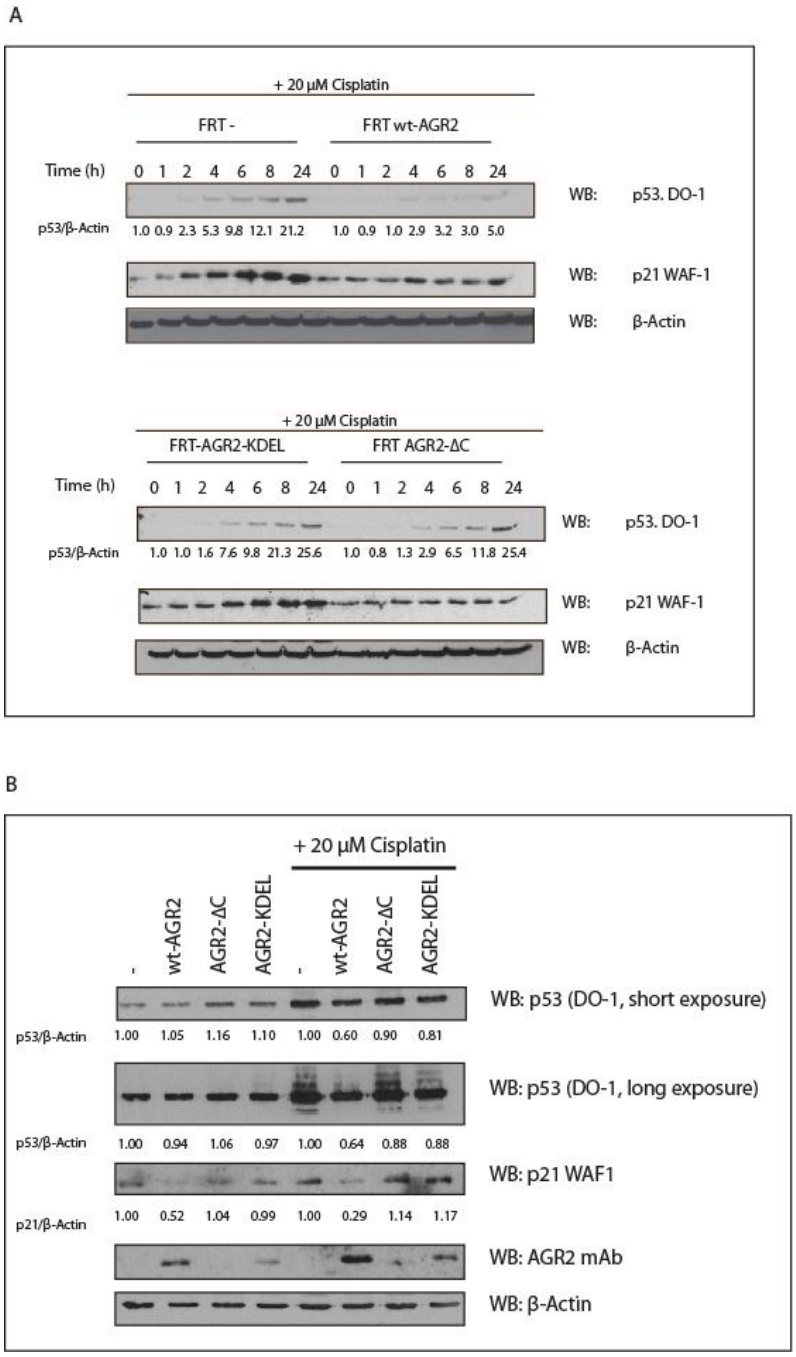


Figure 4-6 Evaluation of p53 levels of the isogenic A375-FRT cell lines expressing wild-type AGR2, AGR2-KDEL and AGR2-ΔC. (A) 24 hour time course of 20 µM cisplatin-induced DNA damage, with western blot detection of p53 levels, p21 activation and β-actin for normalisation. Densitometry quantitation was carried out using ImageJ, relative to β-actin. (B) End-point comparison of 24 hour no treatment or 20 µM cisplatin treatment of isogenic cell panel of AGR2-null, wild-type AGR2, AGR2-ΔC and AGR2-KDEL expressing cells, probed for p53 levels with both a short and long exposure, p21-activation and β-actin for normalisation. Densitometry quantitation was carried out using ImageJ, relative to β-actin.

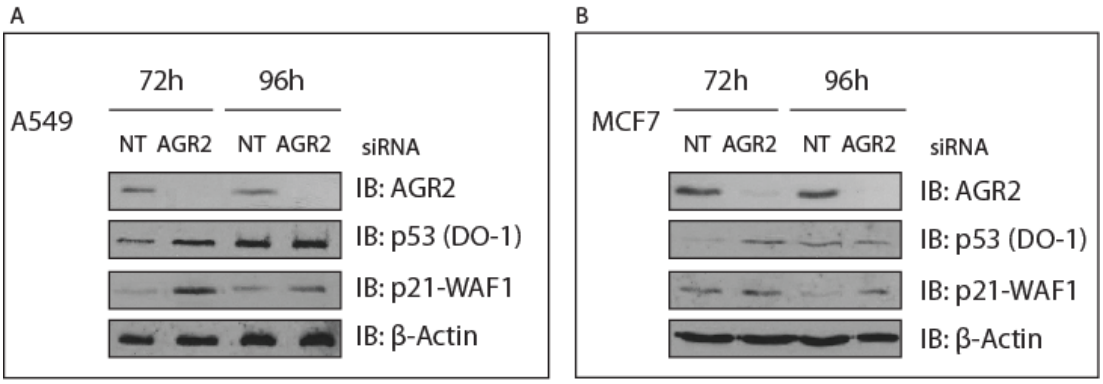


Figure 4-7 siRNA mediated AGR2 silencing selected cancer cell lines. (A) A549 and (B) MCF7 cancer cell line models and the effect of protein depletion on p53 expression and p53-dependent activity of p21 at 72 hours and 96 hours post-siRNA transfection. siRNA was targeted specifically to AGR2, or a non-targeting nonsense control sequence was used.

4.2.4 Data-driven analysis of SILAC expression ratios of wt-AGR2 expressing and AGR2-null isogenic cell.

Further to the global bioinformatical approach to the effect of wt-AGR2 expression, a data-driven methodology was applied to extract as much information from the data set as possible. Significant expression changes (±20%) from the parental cell line initially were ranked and the most altered expression levels identified (Table 4-3 and 4-4). The first application of this ranked data was to identify if any published AGR2-interacting proteins were detected in the proteomic screen and of these if any were significantly misregulated to provide an insight into the oncoprotein functions of AGR2. Of the known AGR2 interactome, only five partners were identified and quantified by mass spectrometry, these being α-dystroglycan (172), CD59 (the human orthologue of newt Prod-1) (154;155), HECTD1 E3 ubiquitin ligase (169), Mucin-18 (165;233) and RUVBL2 (Reptin) (199) (Figure 4-8A). Of these only Mucin-18 reached the significance of 20% upregulated (log_2 of 1.2-fold up-regulation = 0.263), dystroglycan and HECTD1 were significantly down-regulated (below a log_2 = -0.322). Only dystroglycan has a well validated role in metastatic cancer (335) and AGR2 interaction (172) but it's down regulation does not indicate a clear pro-growth effect of AGR2 expression. The interaction of AGR2 and HECTD1 has been identified by yeast-2-hybrid yet requires further on-going validation before the interaction can confidently studied. In addition, the up-regulation of mucins are not surprising due to the molecular chaperone function attributed to AGR2, particularly in the disulphide shuffling of cysteine rich mucin glycoproteins (165;233). This provides an internal control, suggesting that AGR2 is performing a previously published function of the native protein in our experimental isogenic cells.

These ranked data were subsequently subject to literature review, and compared to the ranked expression changes of AGR2-ΔC expressing cells, in order to highlight individual protein expression changes which might influence the observed characteristics of wt-AGR2 function in growth promotion, ER stress response and p53 suppression. Two proteins identified whose inferred function may give an insight into how wt-AGR2 functions as an oncoprotein, were the two proteins which

exhibited greatest upregulation following wt-AGR2 introduction, tumour susceptibility gene 101 (TSG101), and Ki-67, which demonstrated a 3.5-fold and 3.1-fold increase in expression respectively (Table 4-3, Figure 4-8B and C).

Ki-67 is a nuclear and nucleolar protein, thought to be necessary for cellular proliferation and is commonly used in immunohistochemical studies as a marker of proliferative growth in clinical tissue (336). Ki-67 is present in most cell types and is detectable during all active stages of the cell cycle, but absent in resting (G_0) cells. An increase in Ki-67 would suggest that more cell proliferation is taking place in this condition, and this may be coupled to the previously identified p53-activity attenuation (Figure 4-8B). Subsequently, validation of the proteomic data analysis was sought, to assess whether Ki-67 was reproducibly upregulated in wt-AGR2 expressing cells. Due to the large molecular weight of the Ki-67 protein (359 kDa), it was decided immunoblot transfer would not be suitable due to inaccuracies reliant on the charge mediated transfer of protein to nitrocellulose. Thus a fluorescence activated cell sorting (FACS) methodology was utilised in order to determine whether the differential expression of Ki-67 between conditions could be visualised. Ki67 expression is indicative of cell proliferation, therefore a comparison was made of the proportion of Ki67-positive cells in the experimental population. Subsequently, FACS analysis confirmed that compared to AGR2-null A375 cells, the wt-AGR2 expressing cell demonstrated an increased expression of Ki-67 (Figure 4-9). These results validate Ki-67 as an upregulated effector of AGR2 signalling, and are consistent with the enhanced migration activity linked to AGR2 (Figure 3-6, Figure 3-9) (164;188;270). The knowledge of AGR2-induced proliferative activity, validated by Ki-67, linked to the p53-suppression drives further interrogation of the ranked proteomic data to identify co-factors which may mediate the AGR2-dependent down-regulation of p53. Studies of Ki-67 acknowledge the challenging puzzle of deciphering Ki-67 function, and it's up-regulation following wild-type AGR2 expression is another attribute of this multifaceted protein whose role is clearly essential in cell growth (337).

Table 4-3 Ranked SILAC expression data presenting the ten most up-regulated protein expression changes as a result of wild-type AGR2 or AGR2-ΔC gene introduction.

10 most upregulated protein expression

Wild-type AGR2 Expressing Cells

Protein IDs	Protein Descriptions	*Normalised Relative Expression Values*	*Log2 Ratios*	Uniprot	Peptides	Sequence Coverage [%]
IPI00018434	Isoform 1 of Tumor susceptibility gene 101 protein	***3.5344***	***1.8215***	Q99816-1	2	4.4
IPI00004233	Isoform Long of Antigen KI-67	***3.1386***	***1.6501***	P46013-1	3	2.8
IPI00002335	Huntingtin	***2.7066***	***1.4365***	P42858	3	1.9
IPI00941193	48 kDa protein	***2.5402***	***1.3449***	A3F719	4	12.2
IPI00743894	Isoform 1 of Zinc finger C3H1 domain-containing protein	***2.2781***	***1.1878***	O60293-1	3	3.6
IPI00747462	cDNA FLJ54004, moderately similar to Band 4.1-like protein 2	***2.2355***	***1.1606***	B4DJ76	9	43.9
IPI00295772	cytochrome P450, family 51, subfamily A, polypeptide 1 isoform 1	***1.9671***	***0.9761***	Q16850-1	4	14.7
IPI00550882	Pyrroline-5-carboxylate reductase	***1.9460***	***0.9605***	B4DMU0	3	10.4
IPI00844014	Uncharacterized protein C9orf114	***1.8037***	***0.8510***	Q5T280	2	9.6
IPI00018196	Notchless protein homolog 1	***1.7595***	***0.8151***	Q9NVX2	1	2.3

AGR2-ΔC Expressing Cells

Protein IDs	Protein Descriptions	*Normalised Relative Expression Values*	*Log2 Ratios*	Uniprot	Peptides	Sequence Coverage [%]
IPI00941193	48 kDa protein	3.899543	1.963305	A3F719	4	12.2
IPI00890779	Isoform 13 of Dysferlin	2.877175	1.524653	O75923-13	31	20.8
IPI00935722	OTU domain-containing protein 6B	2.676563	1.420381	Q8N6M0	1	5.8
IPI00845373	Isoform 1 of Nuclear factor NF-kappa-B p100 subunit	2.303684	1.203943	Q00653-1	12	19.9
IPI00409607	Isoform 1 of Calmodulin-regulated spectrin-associated protein 2	2.210841	1.144595	Q08AD1-1	2	2.4
IPI00304232	Ribosome biogenesis protein WDR12	2.180031	1.124348	Q9GZL7	2	7.1
IPI00294578	Isoform 1 of Protein-glutamine gamma-glutamyltransferase 2	2.167843	1.11626	P21980-1	29	51.8
IPI00101374	cDNA FLJ61658, highly similar to Transmembrane 9 superfamily protein member 1	2.140867	1.098195	O15321	1	1.5
IPI00031768	Protein Hook homolog 3	2.094993	1.066945	Q86VS8	2	3.2
IPI00004472	Isoform 1 of Serine/threonine-protein kinase WNK1	2.044053	1.031432	Q9H4A3-1	6	3.1

Table 4-4 Ranked SILAC expression data presenting the ten most down-regulated protein expression changes as a result of wild-type AGR2 or AGR2-ΔC gene introduction.

10 most downregulated protein expression

Wild-type AGR2 Expressing Cells

Protein IDs	Protein Descriptions	*Normalised Relative Expression Values*	*Log2 Ratios*	Uniprot	Peptides	Sequence Coverage [%]
IPI00021634	Kinesin light chain 2	***0.0313***	***-4.9992***	Q9H0B6	6	17.4
IPI00554525	NHS protein	***0.0516***	***-4.2773***	Q6T4R5-2	1	1
IPI00027280	Isoform Beta-2 of DNA topoisomerase 2-beta	***0.0725***	***-3.7851***	Q02880-1	2	2.8
IPI00027232	Insulin-like growth factor 1 receptor	***0.2203***	***-2.1825***	P08069	3	6.7
IPI00409601	Isoform 1 of Ral GTPase-activating protein beta subunit	***0.2573***	***-1.9585***	Q86X10-1	3	4.7
IPI00296388	Isoform 2 of Bromodomain adjacent to zinc finger domain protein 2A	***0.2918***	***-1.7772***	Q9UIF9-1	1	1.5
IPI00032038	Isoform 1 of Carnitine O-palmitoyltransferase 1, liver isoform	***0.3068***	***-1.7046***	P50416-1	4	8.7
IPI00386208	Gastric-associated differentially-expressed protein YA61P	***0.3234***	***-1.6285***	Q9NZ23	8	80.9
IPI00301139	Isoform 1 of Mediator of RNA polymerase II transcription subunit 17	***0.3520***	***-1.5064***	Q9NVC6-1	3	8.6
IPI00413451	Putative uncharacterized protein DKFZp686I04222	***0.3679***	***-1.4427***	B2RBA8	2	6.4

AGR2-ΔC Expressing Cells

Protein IDs	Protein Descriptions	*Normalised Relative Expression Values*	*Log2 Ratios*	Uniprot	Peptides	Sequence Coverage [%]
IPI00027280	Isoform Beta-2 of DNA topoisomerase 2-beta	0.129716	-2.94657	Q02880-1	2	2.8
IPI00072377	Isoform 1 of Protein SET	0.196326	-2.34868	Q01105-1	8	41
IPI00554525	NHS protein	0.22967	-2.12237	Q6T4R5-2	1	1
IPI00002335	Huntingtin	0.235531	-2.08601	P42858	3	1.9
IPI00410034	Isoform 1 of Sodium-coupled neutral amino acid transporter 2	0.238448	-2.06826	Q96QD8-1	5	16.6
IPI00217468	Histone H1.5	0.254236	-1.97576	P16401	3	16.4
IPI00012545	Isoform TGN51 of Trans-Golgi network integral membrane protein 2	0.277091	-1.85157	O43493-1	5	13.3
IPI00644502	Isoform 1 of GTPase-activating Rap/Ran-GAP domain-like protein 3	0.2878	-1.79686	Q5VVW2-1	2	2
IPI00026154	cDNA FLJ59211, highly similar to Glucosidase 2 subunit beta	0.311942	-1.68065	B4DJQ5	29	63.4
IPI00333015	Isoform 2 of Spectrin beta chain, brain 1	0.32176	-1.63594	Q01082-3	140	74.3

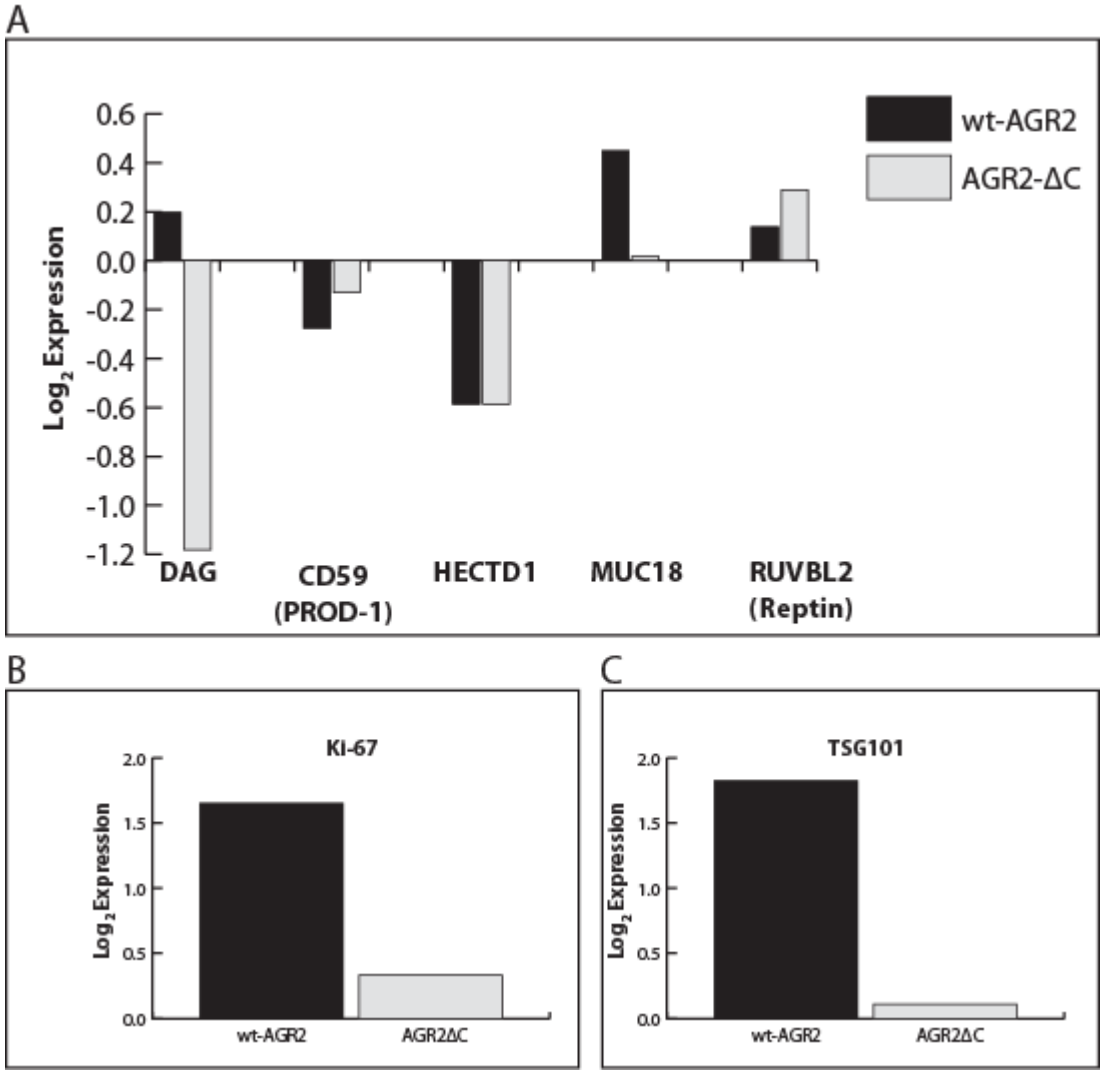


Figure 4-8 Expression changes of highlighted proteins as a result of wt-AGR2 or AGR2-ΔC gene recombination. (A) Published interacting partners of AGR2 log2 expression changes as a result of experimental gene introduction. (B) Ki-67 demonstrates significant upregulation, the second most increased, following wt-AGR2 gene expression, but is not induced by AGR2-ΔC. (C) TSG101 is the most upregulated protein following wt-AGR2 expression compared to AGR2-null isogenic cells; this is specific to the wild-type protein as TSG101 is not upregulated by AGR2-ΔC expression.

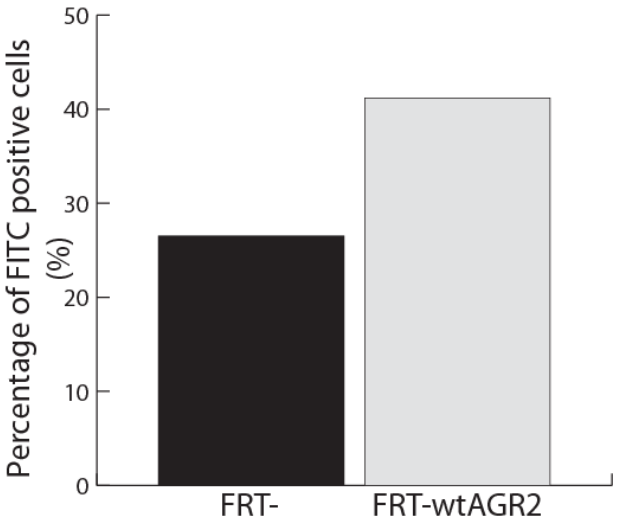


Figure 4-9 Validation of Ki-67 as a protein upregulated by wt-AGR2. Fluorescence activated cell sorting (FACS) analysis of AGR2-null parental A375 FRT- cells with A375 FRT-wtAGR2 expressing cells using a FITC conjugated anti-Ki-67 antibody. FITC reactivity was detected at 488nm and presented as a summary of the percentage of cells expressing Ki-67 protein.

4.2.5 Identification and validation of TSG101 as an effector in the suppression of p53 dependent activity by wt-AGR2 expressing cells

In order to define a mechanism to explain how AGR2 can suppress p53 protein, we took advantage of the fact that Ingenuity Pathway Analysis identified components of the TP53 pathway (Table 4-1) and subsequent validation (Figure 4-6) collectively with the ranked most upregulated protein expression was identified as the previously identified p53-inhibitor TSG101 (Table 4-3) (338). Defects in TSG101 have previously been linked to multiple cancers (339;340) and steady state gene expression of this gene product has been described as necessary for genome stability and cell cycle regulation (341;342). TSG101 forms part of the endosomal sorting complex required for transport-1 (ESCRT-1), and functions as a regulator of vesicular trafficking, whose linkage to AGR2 might provide an insight for the role of AGR2 as a PDI in protein maturation and protein trafficking (169). Intriguingly, TSG101 has previously been identified as participating in the well studied p53-MDM2 autoregulatory loop, by affecting protein decay (338). The Ubc domain of TSG101 disrupts the ubiquitination of MDM2, thus inhibiting MDM2 decay and elevating its steady state level, and that these events are associated with the down regulation of the p53 protein.

Subsequent biochemical validation of the upregulation of TSG101 in wt-AGR2 expressing cells, and its role in influencing p53 protein levels was performed. By Western blot, basal TSG101 levels are significantly greater in wt-AGR2 expressing cells, compared to the isogenic AGR2-null cells (Figure 4-10A), consistent with the conclusions drawn from the SILAC quantitation data. Due to the cells only differing on the expression of the wt-AGR2 gene, we can thus conclude that the upregulation is a direct result of AGR2 protein presence. Further, transient transfection of TSG101 (using TSG101-pCMV-CL5 expression plasmid) into A375-FRT-wtAGR2 expressing melanoma cells suppressed further endogenous p53, and p53-dependent activity, particularly when proteosomal degradation was blocked using MG132 (Figure 4-10B). Thereby, supporting the hypothesis that the overexpression of TSG101, as a result of AGR2 presence, was having an attenuation effect on the expression of p53, and p53-dependent proteins. The blocking of proteosomal

degradation in this experiment ensured that the difference in p53 levels could be visualised as in non-MG132 treated samples; the endogenous p53 was clearly being rapidly turned over. For confirmation of the p53-suppressive effect of TSG101, exogenous TSG101 was titrated into two wild-type p53 cell lines, A375 (AGR2-negative) and MCF7 (AGR2-positive) and the effect observed (Figure 4-10C). As expected even very small changes in the expression of TSG101 attenuated p53 expression (342), with as little as 0.4 μg per 2 mL culture, suppressing p53, and concurrently p21 levels. This experiment indicated, that AGR2 was not a necessary cofactor for the p53-suppressive activity of TSG101, in fact, in AGR2-null A375 cells, TSG101 suppressed p53 independently of AGR2 expression.

The p53-suppressive nature of TSG101 was further probed using cisplatin-induced DNA damage, and siRNA mediated protein expression silencing in authentic AGR2-expressing cancer cell lines. A549 cells have a wild-type p53 pathway and an active TGF-β dependent pathway that regulates AGR2 expression in an oestrogen-independent manner (180). When these cells were transfected with TSG101, whether cells were damaged or not, p53 levels are suppressed (Figure 4-11A, lanes 2, 4, 6 and 8 vs 1, 3, 5 and 7). When A549 cells were damaged with cisplatin, there was still a remaining dominant effect of TSG101 over p53 protein levels (Figure 4-11A, lanes 6 vs 5). When MG132 was added to cells, the effect of TSG101 on p53 protein turnover can still be seen in the presence of cisplatin (Figure 4-11A lane 8 versus 7). These data show that in this cell line, forced expression of TSG101 can suppress p53 protein induction and/or increase p53 protein turnover under a set of experimental conditions including the basal state or cisplatin damage. We also evaluated the TSG101 effects in MCF7 cells, which also have a wt-p53 pathway but the expression of AGR2 protein in this cell line is oestrogen dependent. Similar to A549 cells, under each condition, ectopic expression of TSG101 suppressed p53 protein levels (Figure 4-11B, lanes 2, 4, 6 and 8 compared to 1, 3, 5 and 7).

To complement the experiments above that evaluate how ectopic expression of TSG101 can impact on cells endogenously expressing both wt-p53 and AGR2, we used a targeted siRNA approach to determine whether reduction in endogenous TSG101 impacted on p53 protein turnover. In A549 cell, depletion of TSG101 using

siRNA only had a significant impact on stabilising p53 protein levels in the presence of MG132 (Figure 4-10C lanes 5 and 6 vs 1-4), which is consistent with the effect of transfected TSG101 on p53 levels in A375-FRT-wtAGR2 cells (Figure 4-10B). Similarly, in MCF7 cells, TSG101 depletion most impacted on p53 protein in the presence of MG132 (Figure 4-11D lanes 5 and 6 vs 1-4), although an effect on p21 protein was observed under all conditions where TSG101 was silenced. Together, these data suggest that the over-production of TSG101 induced by the reprogrammed A375-FRT-wtAGR2 cell can, in part, relate to the ability of TSG101 to stimulate the turnover of p53 in a range of cell types.

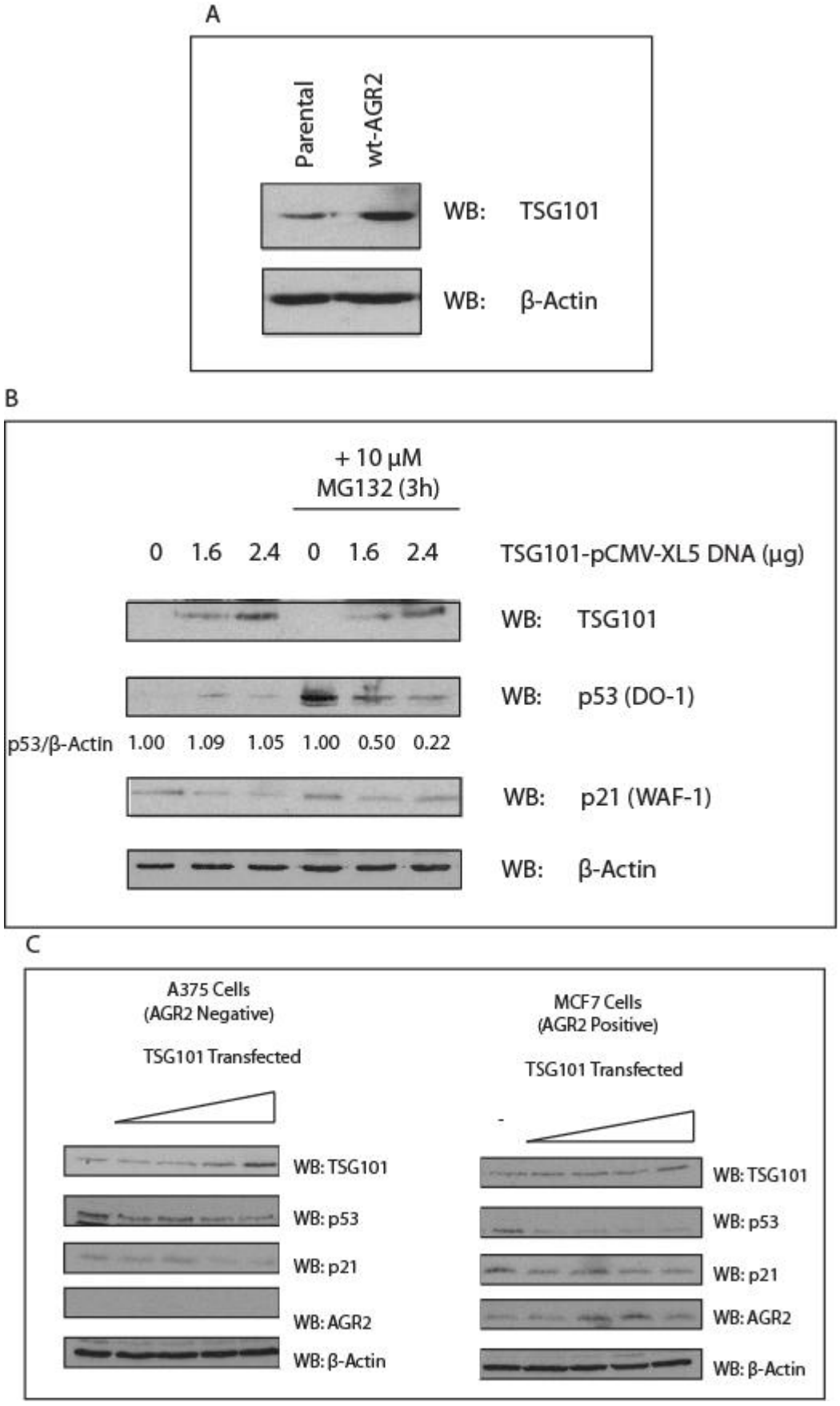

A
Parental
wt-AGR2
WB: TSG101
WB: β-Actin
B
+ 10 μM
MG132 (3h)
0 1.6 2.4 0 1.6 2.4
TSG101-pCMV-XL5 DNA (μg)
WB: TSG101
WB: p53 (DO-1)
p53/β-Actin 1.00 1.09 1.05 1.00 0.50 0.22
WB: p21 (WAF-1)
WB: β-Actin
C
A375 Cells
(AGR2 Negative)
TSG101 Transfected
MCF7 Cells
(AGR2 Positive)
TSG101 Transfected
WB: TSG101
WB: p53
WB: p21
WB: AGR2
WB: β-Actin
WB: TSG101
WB: p53
WB: p21
WB: AGR2
WB: β-Actin

Figure 4-10 The effect of TSG101 expression on p53 protein level. (A) Basal TSG101is upregulated in engineered A375-FRT-wt-AGR2 expressing cells relative to the parental cell line, consistent with SILAC data from Table 4-3. (B) Transfection of increasing amounts of TSG101-pCMV-XL5 expression plasmid for 18 hours incubation then followed by 3 hour incubation with 10 μM MG132 to determine whether TSG101 affects p53 turnover in A375 cells. Lysates were blotted with antibodies against TSG101, p53, p21 and a β-actin loading control. (C) TSG101 was titrated into transfection of A375 (AGR2-null) and MCF7 (AGR2-positive), wild-type p53 expressing cells, 0, 0.4, 0.8, 1.2, 1.6 and 2.0 μg expression plasmid DNA per 2 mL 70% culture of cancer cells. Western blot was probed with antibodies as indicated.

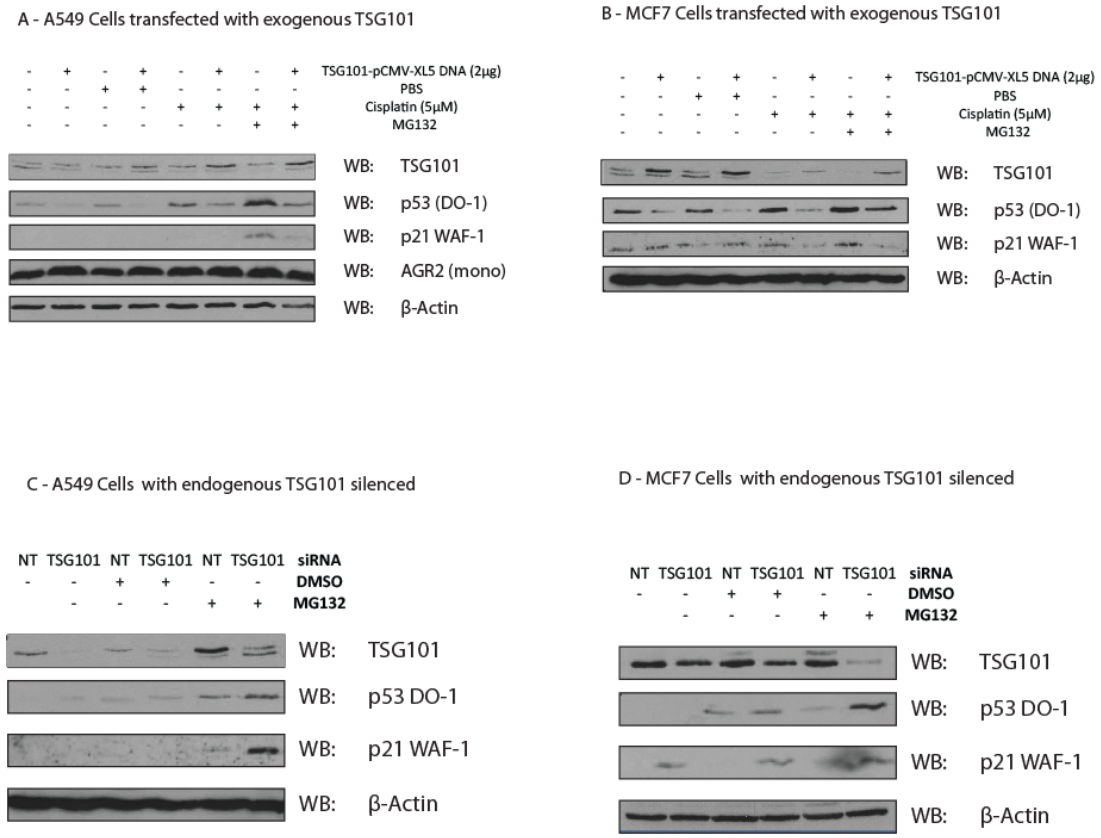


Figure 4-11 The effects of TSG101 expression on p53 protein levels. (A) and (B) TSG101 was transfected into endogenously expressing AGR2 cells (as indicated A549 or MCF7) and followed by addition of carrier (PBS), cisplatin, and MG132, as indicated, to determine whether TSG101 suppresses p53 protein in response to cisplatin induced DNA damage. (C) and (D) Targeted siRNA was transfected into cells to determine whether depletion of TSG101 in endogenously expressing AGR2 cells altered the turnover of p53 protein, as defined by p53 protein levels without or with MG132 treatment. Lysates were blotted with antibodies to p53, p21, TSG101 and a β-actin loading control.

4.3 Discussion

The aberrant expression of Anterior Gradient-2 is correlated with several human pathologies including cancer (169). The mechanism of the over-expression of AGR2 in diseased tissue is not clearly understood, with studies highlighting possible contribution in endoplasmic reticulum stress response as a molecular chaperone (179), misregulated EGF signalling (194) and p53-dependent activity inhibition (150;187) driving oncogenic signals in cancer cells, and cell proliferation (164;231;244) and cell cycle signalling pathways (231;262). However, co-factors which might mediate these functions are poorly understood. Initial studies on the molecular basis of AGR2 signalling comes primarily from protein-protein interaction studies, via yeast-2-hybrid and co-immunoprecipitation experiments which have implicated C4.4A, α-dystroglycan (172) and Reptin (199), along with other protein interaction hits which remain to be full validated (169). These studies have not yet elucidated a clear signal transduction pathway, or co-factors involved, to explain the survival benefit of AGR2 over-expression. Further, in oncology, AGR2 has been implicated in drug resistant tumours against tamoxifen (174;175) and cisplatin (248), as a pro-metastatic factor (164;218), as well as disrupting homeostatic signalling pathways (150;194) and remains under study in cancer drug discovery programmes (169).

To drive our understanding of the effects of AGR2 overexpression, in Chapter 3, we engineered and characterised an isogenic cell panel constitutively expressing our gene of interest. These cells provided an appropriate tool for the study of AGR2 signalling, since they differ from control cell lines by only a single variable, the expression of the AGR2 gene (or mutant AGR2 gene). In order to utilise this experimental tool aptly to give the greatest opportunity of deciphering the molecular function of AGR2, we took a systems biology approach to the analysis of the cell lines using transcriptomics (Chapter 3) and proteomics, described herein. The whole genome gene expression analysis had given little insight into the function of AGR2, leading to the assumption that AGR2 was not driving a direct effect on the

transcription of genes or that any effect at the mRNA level of the cell was slight, since as highlighted, the cells exhibited a modest level of AGR2 expression. Since AGR2 is predominantly an ER-resident PDI (165;179), it was considered whether its chaperone function might have a primary effect on the maturation of endogenous proteins. This was justified since the low level AGR2 expression in the engineered A375-FRT-wt-AGR2 cells had previously been identified to supplement the homeostatic CHOP response to tunicamycin induced endoplasmic reticulum stress (Figure 3-5C). As such, we evaluated whether the addition of the single AGR2-expressing allele could reprogram the cellular steady-state proteome under conditions where changes in the transcriptome were negligible.

In order to unravel the mechanism whereby AGR2 gene expression affects cell function, a quantitative proteomics approach was sought. Progress in mass spectrometry instruments and data analysis packages mean that there are now several different technologies which can be implemented to quantify the protein expression landscape of the experimental condition (Table 4-5). Protein quantitation can be broadly split into two categories, those which measure (i) relative quantitation and (ii) absolute quantitation. Relative quantitation methodologies compare two or more samples to one another, e.g. to a control, to provide a fold change of the experimental condition to the untreated sample. In its simplest form, 2-dimensional gel electrophoresis coupled to fluorescent dyes (as mentioned previously) allow protein levels to be quantified using the intensity of the incorporated fluorescent dye relative to the same spot from a control lysate. In MS-based proteomics, the reliance is based on stable isotope dilution theory, which states that a stable isotope-labelled peptide is chemically identical to its native counterpart; therefore the two peptides also behave identically during chromatographic and mass spectrometric analysis. The sensitivity of a MS detector means that the mass difference between the labelled and unlabelled isoforms of the peptide can be differentiated, and quantification can be achieved by comparing, relatively, their signal intensities. Isotope labels can be introduced into the sample in several ways: (i) metabolically, (ii) through chemical reactivity, (iii) enzymatically or (iv) as an external standard of spiked synthetic peptide to normalise to (343). Recently, label free strategies have begun to emerge, such as spectral

counting where the number of spectra acquired of a specific peptide/protein are used as an estimate of abundance in the sample.

The earliest possible point for protein labelling is during nascent protein synthesis, and thus at the metabolic level, allows for the most accurate quantitative MS method (330). The SILAC method used in this study requires the metabolic labelling of cells in tissue culture (324), in conditioned media incorporating isoforms of amino acids with heavy elements allowing labelling of all nascent proteins in the experimental condition. Following specific trypsin cleavage, every enzymatic product, except for the C-terminal most peptide, incorporates an isotopic label. This method results in a mass differences between otherwise identical precursor peptides which can be detected, grouped and the spectral peak integrated to provide a measure of quantitation. Some level of multiplexing is possible, limited to three different growth media permutations, however up to five is possible in experiments that have at least one common sample allowing normalisation (287;344;345). The key advantage of SILAC is that cells from various conditions can be combined as early as at the level of intact cells and subject to concurrent processing. Thereby removing sources of quantification error introduced by biochemical or MS procedures as all experimental protein populations will be treated in the same manner (330). Metabolic labelling was, until recently, limited to cells grown in tissue culture, however insights are beginning to overcome the economical and practical constraints of labelling whole organisms and *in vivo* labelling of *Drosophila* (346), rat (347), mice (348) and nematodes (349) are reported.

Comparatively, proteins, or peptides, from lysates can be labelled post-biosynthesis by chemical or enzymatical derivation strategies. Following proteolytic digestion with trypsin or Glu-C, ^{18}O can be enzymatically incorporated into the C-terminus of peptides of one of the experimental conditions. Utilising the heavy water in the trypsin digestion step, one or two heavy oxygen atoms are incorporated into the tryptic peptide (350). As a result, a mass shift of 2 Da per oxygen atom incidence can be identified (351). The challenged of ^{18}O labelling is that full labelling is difficult, and different peptides have different affinities for the ^{18}O tag resulting in a complex analysis. Proteins and peptides can be chemically labelled also, ICAT

(isotope-coded affinity tag (352)) is one example which labels cysteine side chains with deuterium atoms and biotin groups allowing differential quantitation. Cysteine being a rare amino acid means that complex samples are reduced in complexity; however, for proteins lacking in cysteine residues the technique has much reduced relevance. A further group of chemical labels target the N-terminal of the peptide and the epsilon-amino group of lysine residues, utilising specific NHS chemistry to conjugate the isotope tags to the peptide being isotope tags for relative and absolute quantification (iTRAQ) (353) and tandem mass tags (TMT) (354). These concepts introduce isobaric tags to label peptides which do not influence chromatographic nature and can only be differentiated by MS. The ratio of detection of the tags can then be used as a quantitation measure to detect the difference between the levels of proteins from different samples. iTRAQ and TMT allow significant multiplexing of the experimental conditions of up to 8 samples, which is particularly useful in studies of time courses or in providing biological replicates. The chemical labelling step however, adds an additional processing step between samples, and thus introduces a level of error into the experiment. Label-free methods can avoid the problems caused by the labelling steps. Label-free techniques however require complex algorithms to normalise the quantitation results based on the length of the protein as longer proteins have a greater chance of the detection of multiple peptides (355). Currently, label-free strategies lack the accuracy of the labelled counterparts yet continued development of methodologies is proving fruitful (356). Absolute quantitation (AQUA) uses heavy isotopic peptides of a known protein at standardised concentration spiked into the sample and compared to the detected protein, allowing determination of how much of that specific protein is present in the sample (357). One drawback of this method is that you require knowing which protein you want to quantify initially, thus can quantify the abundance of a known biomarker, but cannot be used for biomarker discovery.

Table 4-5 Summary of quantitative proteomics strategies

Methodology	Labelling Strategy	Relative/Absolute	Advantages	Caveats
2-DIGE	Spot analysis or Fluorescent label	Relative	Simple, measures whole protein	Limited range of quantitation
iTRAQ / TMT	Isobaric tags	Relative	Allows multiplexing	Additional chemical modification step
SILAC	Heavy isotope metabolic tag	Relative	Concurrent processing of samples	Requires propagation in tissue culture. Limited *in vivo* use
^{18}O Labelling	Heavy oxygen	Relative	Label during trypsin digest	Complete labelling is rare
Spectral Counting	None	Relative	No labelling	Complex data analysis. Reproducibility issues
AQUA	Spiked heavy peptide from known protein	Absolute	Absolute quantitation	Only suitable for known targets

The triple labelled SILAC methodology provided the most appropriate platform for this study as it not only allowed comparison of AGR2-null and AGR2-expressing cells, but also the comparison with the AGR2-ΔC mislocalised gene product to control for intracellular specificity. SILAC quantitation has proven effective in numerous publications including a comparable study of engineered isogenic cell lines expressing mutant phosphoinositide3-kinase (PI3K) (332). However SILAC does require cells to be propagated in heavy medium for several passages before the vast majority of the proteins in the sample are appropriately labelled and thus able to be separated by MS. It is essential that all proteins are metabolically labelled with the isotope tag as even a small percentage of unlabelled amino acid in the 'labelled' population is detrimental to the experiment as this contributes to the unlabelled signal, thereby introducing quantitation errors. Five cell doublings are the generally accepted minimum to allow incorporation of the heavy amino acid, as ~97% of proteins should be in the heavy state (287), however it is essential to check that there are no unlabelled peaks in the labelled samples (data not shown, carried out by R. Lenobel, Palacky University, Olomouc, CZ and L. Hernychova, Masaryk Memorial Cancer Institute, Brno, CZ) prior to mixing and SILAC analysis of the samples. This

allowed confidence in the quantitation of normalised expression ratios between the labelled and unlabeled conditions. It is worth noting that due to the metabolic nature of SILAC (358), it can also be used for quantifying of protein turnover experiments, through a spike in of isotope containing media and measurement and analysis of the proteome at time points thereafter. This provides a detection method for the change in the incorporation rate of SILAC-labelled amino acids in specific proteins, providing a dynamic measure of protein flux in the cellular proteome. For future experiments, it may be relevant to interrogate the effect of AGR2 expression on protein flux, however this study presents initial attempts to decipher the steady-state effect of AGR2 on protein signalling and thus full labelling was required.

The choice of the labelling strategy of SILAC amino acids was also taken into consideration. With several possible isotopically labelled elements commercially available including ^{13}C, ^{15}N and ^{2}H multiple options for labelling are available. The SILAC amino acids used for labelling should also be essential for the survival of cultured cells, ensuring that the amino acid supplemented into the tissue culture media is the only source of that particular amino acid for the growing cells. Leucine (324), lysine (359) and methionine (360) are essential amino acids commonly used for SILAC labelling. Arginine is not strictly an essential amino acid, rather is conditionally essential and is mostly obtained through the diet in whole organisms, it has been described as essential in many cultured cell lines (361;362) and has previously been used successfully in SILAC labelling (345;363). The most important criterion of selecting isotope labelled amino acids, as described by the pioneers of the technique (287), is that they should provide a mass difference of at least 4 Da from the normal isotope abundance amino acid to avoid overlapping of isotopic distributions. Therefore, we utilised a medium label of ^{13}C arginine and ^{2}H lysine (R6K4) and a heavy label of and ^{13}C and ^{15}N lysine (R10K8), which met the desired 4 Da separation principle. It should be acknowledged that due to the formulation of tissue culture media, such as DMEM, that amino acids and other nutrients are generally provided in excess. This highlights an issue of the labelling strategy used, coupled to the molecular basis of arginine that in conditions of excess arginine, an arginine to proline interconversion is known to take place. This may

have a downstream influence on protein quantitation accuracy as highlighted by a study of HeLa and HEK293 cells (363). As a result, two distinct peak clusters for all proline containing peptides would be identified in the labelled state, however, this was not found to be an issue in this study. If a two-peak cluster was identifiable then the arginine concentration of the labelling media could be empirically titrated until the proline peak did not indicate two species.

Two further considerations were required to ensure accuracy in SILAC quantitation by minimising the abundance of natural abundance amino acids in the 'medium' and 'heavy' labelled metabolic labelling conditions. Firstly, through the use of dialysed serum (324), depleting the media of bioactive peptides below 10 kDa molecular weight cut off. Thereby removing small molecules in the growth supplement containing any instance of the normal amino acid residue which may be metabolised and used to mis-label nascent peptides. This step is essential to the accurate quantitation of protein expression levels; however, it is widely known that some cells do respond anomalously to dialysed sera lacking specific low molecular weight factors required for attachment and growth for example. Fortunately, A375-FRT cells did not show any detectable phenotypic response to dialysed serum, there was no change in cell morphology, cell proliferation or attachment to the substratum (data not shown), thus dialysed serum was appropriate (if necessary publications have highlighted that purified growth factors or a small percentage of normal serum can be supplemented into the media (364), however this must be acknowledged as a source of potential quantitation error). The second, though generally not followed as strictly as other avoidances of unlabelled amino acids, is the use of a PBS-based cell dissociation buffer ensures that no external protein is introduced into the experimental tissue culture. Normal 0.05% Trypsin-EDTA enzymatically cleaves adhesion proteins allowing detachment of cells from the substrate for propagation; however the trypsin enzyme provides a source of normal amino acids which might remain in the medium allowing uptake and integration into synthesising proteins. Theoretically, it may be sufficient to wash cell pellets with PBS following neutralising of the trypsin reaction before re-plating in the specified media; however,

in this experiment it was optimal to use the PBS-based dissociation buffer for accuracy.

As a consequence of these factors which may influence accurate SILAC-based quantitation, there are several conditions under which the cells must endure without exhibiting significant disruption to allow full proteome labelling, and subsequent precise quantitation. Auspiciously, A375-FRT cell panel did not respond dramatically to tuned growth conditions optimal for SILAC labelling, allowing SILAC to be utilised as the most appropriate method of proteome analysis.

Initial observation of the SILAC ratios of protein expression indicated that the vast majority of proteins quantified in the study did not vary beyond ±20% the expression level of the AGR2-null parental cell line (Figure 4-4). SILAC is the most appropriate method of quantitation for detection of small changes in protein expression (330), and it was clear that wt-AGR2 recombination was driving a specific reprogramming of the cell. Un-biased pathway analysis allows examination of the full scale of proteomic data gathered, without operator prejudice. Due to the vast nature of proteomic data collection, bioinformatics tools provide a key means in the deconvolution of raw data. By submitting the significant expression changes to the IPA software tool, we could identify the main biological pathways affected by over- or under- representation of their constituent parts, and in certain cases predict the activation state of that biological process. In this experiment, IPA identified that the TP53 signalling pathway was most significantly inhibited in the experimental wt-AGR2 expressing cell line, compared to the proteomic profiles of AGR2-null and AGR2-ΔC control cell lines (Table 4-1 and Figure 4-6). Through the modulated expression of 39 proteins linked to p53-signalling function (Table 4-1 and Figure 4-5), the reprogramming of p53-dependent activity by wt-AGR2 was highlighted. Further validation (Figure 4-6) confirmed that the p53-response to DNA damage was attenuated by wt-AGR2 expression and not by any of the subcellular mislocalised mutants; this indicated that the diverged C-terminal motif of AGR2 was indeed necessary for p53-suppressive function, as previously identified for EGFR-activity (188). This complements previous discoveries of the p53 suppressive nature of AGR2 and its potential as a novel drug target for p53 reactivation (150;187;275).

Further, these data show that the synthetically engineered cells can be utilised in unravelling, at least one mechanism of action of AGR2 with regards to it as a proto-oncoprotein. It is intriguing that the reliance on the C-terminal motif for EGF function (188), along with the variety of subcellular distribution of wild-type AGR2 in experimental cells (187), that both C-terminal mutants of AGR2 provided a growth advantage compared to the AGR2-null cells (Figure 3-9), yet neither demonstrated an attenuated p53 response to DNA damage (Figure 4-6). This may highlight a further possible decipherable pro-growth reprogramming in addition to p53 suppression. Other transcriptional pathways disrupted include the E2F1 (activated) and TFEB (suppressed). E2F1 plays a significant role in the regulation of cell proliferation, the cell cycle and is often misregulated in developing tumours (365). Intriguingly, E2F1 is reported to co-operate with p53, via activation of ARF and subsequent inhibition of MDM2, stabilising p53 levels (366). Our data suggest that AGR2 might uncouple this interplay, resulting in p53 levels being attenuated.

It must be considered that proteomic pathway analysis highlights a suppressive effect of wt-AGR2 expression on one of the key homeostatic transcription factors in p53, yet this activity did not correlate to any detectable effect when measured at the transcriptome level (Figure 3-10). Surprisingly, in contrast to the underlying assumption of the flow of information from DNA to phenotype in which perturbations of transcript levels drive protein expression, a lack of correlation between mRNA levels and protein levels are not uncommon (367;368). We must therefore assume that low level AGR2 is functioning at the protein level to sequester p53 expression, and that downstream transcriptional targets are not being affected to such an extent to indicate significance by microarray. Evidence is garnered with the fact that transcripts deemed to be upregulated by the microarray, could not be reproduced by the alternative qPCR method.

Among the proteins whose expression levels were modified by wt-AGR2 (Figure 4-8) includes Mucin-18, a cysteine rich glycoprotein whose expression is increased by 36% following wild-type AGR2 expression. As previously described (165;233), AGR2 participates in the disulphide reshuffling of mucin proteins resulting in an increase in the abundance of mature mucin proteins and a reduction in the

degradation of nascent chains. Known AGR2-interacting proteins RuvBL-2 (Reptin), dystroglycan, and CD59 expressions do not modulate significantly in this report. HECTD1, a putative yeast-2-hybrid identified interaction candidate demonstrates a reduced expression profile, however the HECTD1-AGR2 interaction is still under further study and requires further biochemical validation (169). Ki-67 demonstrates expression >3-fold upregulated compared to the parental cell line; the Ki-67 protein is associated with cell proliferation and used as an IHC marker of active proliferating cells (336). Previous studies of the oncogenic properties of AGR2 make this a relevant over-expressed protein, as it, and subsequent FACS analysis validation of expression levels, indicates wt-AGR2 expressing cells have enhanced expression of a key immunohistochemical marker of proliferation. Coupled to this the observed increase in proliferation and reduction in cell doubling time of wt-AGR2 cells compared to the AGR2-null parental cells (Figure 3-9), we could confirm that AGR2 is indeed, in line with previous publications (164;218;244;249;270) acting as an oncoprotein whose single allele introduction is adequate to present a pro-growth phenotype.

The identification of tumour susceptibility gene 101 (TSG101) as a dominant effect of wt-AGR2 remodelling of the proteome prompted further validation. TSG101 is involved in proteostasis, and protein trafficking due to its involvement in the ESCRT-1 complex of proteins regulating the transport of components of late endosomal compartments (369). TSG101 contains a ubiquitin conjugating (Ubc)-like domain which recognises ubiquitylated substrates and incorporates them into the multivesicular body (370). TSG101 over-expression has recently, somewhat controversially, been implicated in cancer growth, despite initial characterisation as a tumour suppressor gene (371;372). The majority of literature now describes TSG101 as a tumour enhancer in various cancers including breast, ovarian and prostate cancers (340). Steady state levels of TSG101 are maintained within narrow limits (373), and over-expression of TSG101 can lead to neoplastic transformation (374). In complementary studies, *tsg101* silencing in mice reported defective cell proliferation and p53 protein accumulation (together with subsequent $p21^{WAF-1/CIP-1}$ activation) (295), while a reported regulatory loop involving TSG101/MDM2

resulted in MDM2 degradation modulation with a downstream effect on MDM2/p53 signalling (338). Our data also supports the interplay between levels of TSG101 and subsequent effect on p53 levels and activity.

We can therefore summarise that over-expression of AGR2, exhibiting the essential KTEL C-terminal motif, drives an upregulation of two identified co-factors confirming its role as a proto-oncoprotein. Ki-67 is upregulated indicating that cells have an increased proliferative capacity, as previously reported, and the induction of p53-inhibitor TSG101 co-operates to induce p53-dependent activity suppression (Figure 4-12). Further study will therefore begin to interrogate the mechanism whereby AGR2 directly or indirectly upregulates or stabilises TSG101 at the protein level, although the assumption is that AGR2-mediated chaperone maturation is linked to the increased presence of the protein, however this requires further validation through a targeted approach. The complexities lie in unravelling how TSG101, an endosomal sorting complex factor that functions in vesicular trafficking and other pathways, regulate p53, a nuclear transcription factor. The fact that TSG101 overexpression has been highlighted in the disruption of the cell cycle, proliferative capacity and suppression of p53 signalling, coupled to the identification as the outlying, most over-expressed protein of a SILAC proteomics screen of effectors of AGR2 induced cell reprogramming drives the desire to identify the co-expression of AGR2 and TSG101 in clinical tissue arrays to determine whether the coupling of expression of these co-factors is a contributor to tumour development *in vivo*. The data gathered presents a new appreciation of our understanding of AGR2 function in tumour cells, and provides knowledge that may be useful in interrogating AGR2 effectors as a potential therapeutic target.

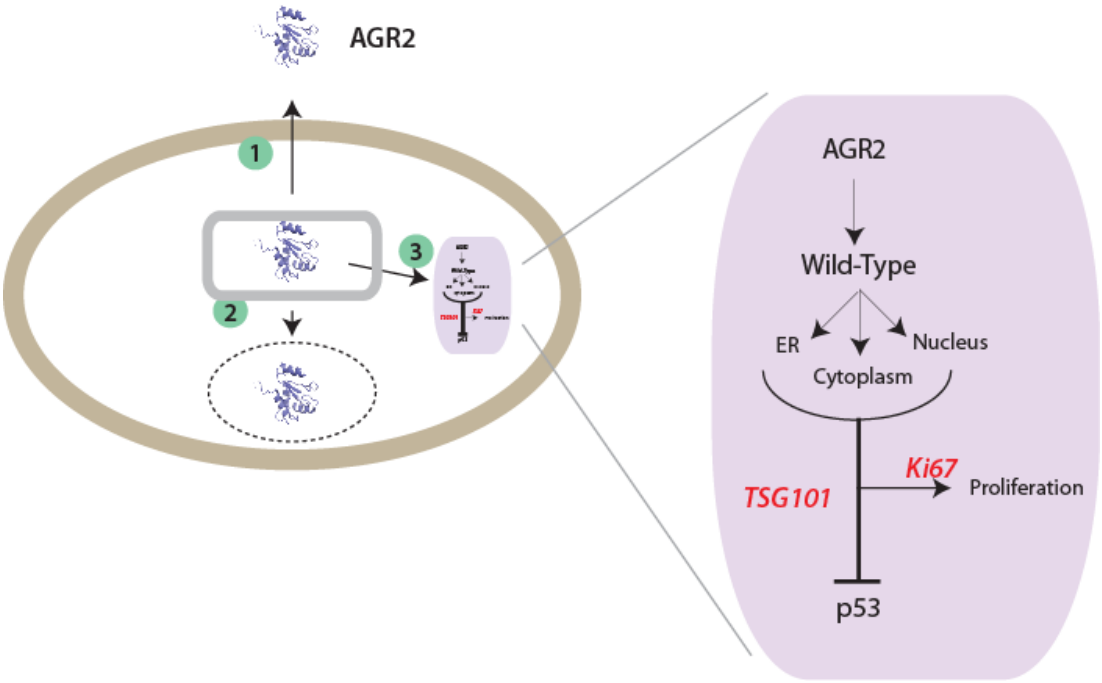


Figure 4-12 Summarised illustration of the dominant effects of engineered AGR2 expression on the cellular proteome. 1- AGR2 is an ER resident PDI that can regulate the secretion of the mucin family of proteins (165;233) as well as pro-oncogenic receptors as defined by yeast-two hybrid, including DAG and C4.4A (172) that might contribute to metastatic properties. 2- A second role outwith the ER is also reported where AGR2 can signal to nuclear transcription pathways to upregulate EGFR receptor pathways (194) and interact with the chaperone protein Reptin (199). 3- The relatively low level AGR2 expression in this report has comparatively minimal effect on the transcriptome of the cell, but SILAC analysis has shown the dominant reprogramming of the proteome that occurs by virtue of AGR2 production. This includes upregulation of the proliferation associated protein Ki-67 and TSG101, the latter of which provides a signalling mechanism that can explain in part how AGR2 can attenuate the specific activity of p53 in response to DNA damage.

Chapter 5: Molecular interrogation of AGR2 confirming a homodimeric structure

5.1 Introduction

The complex properties of the amino acid residues which comprise the building blocks of life, their polymerisation into a linear chain of diverse repeating units, and the interaction with the surrounding environment provide the structural determinants driving the contortion of the polypeptide into the functional protein unit. The protein subunit is stabilised by non-covalent hydrogen bonds, ionic interactions, Van der Waals forces and hydrophobic packing, driving secondary structural motifs, the α-helices and β-sheets. Additionally, intra-molecular non-covalent and covalent interactions determine relationship between interaction partners to form functional, 3-dimensional bioactive complexes (375). This intricate biomolecular organisation drives the final specific spatial conformations, which are essential for the function and activity of the derived protein. The protein structure, therefore, holds the key to expressed domains, motifs and functional folds, which enable sub-categorisation of the molecule into protein families, allowing initial assumptions to be made regarding protein function. In the case of Anterior Gradient-2, structural interrogation can provide us with new information of the possible determinants for functions of AGR2, and indicate characteristics which drive its potential as an oncoprotein. This structural knowledge could then be utilised in a drug discovery programme for compounds which disrupt motifs or interactions essential for driving its growth promoting activity.

As previously described in Chapter 1, AGR2 protein incorporates several structural motifs which have been used to begin to characterise the mature protein function. The N-terminal 20 amino acid residue signal sequence, coupled with the atypical C-terminal endoplasmic reticulum sequence, KTEL (192), and a single CXXS motif has incorporated AGR2 as a distant member of the large protein disulphide

isomerase (PDI) family. This has allowed for the identification of the fact that AGR2 plays a role in the maturation of cysteine rich proteins, functioning as a thiol/disulphide exchange factor implicated in the formation of mixed disulphides in MUC2 (165), MUC1 (180) and MUC5AC (233). However, structure-based analysis of AGR2 has been less clear with regards to the role of AGR2 in abnormal cells, driving cell proliferation (164;231;249), transformation (164) and metastasis (172;208;219;259;270).

The discovery of protein-protein interactions is a key, fundamental goal in life sciences research, and these interactions are emerging as compelling landscapes in the drug discovery field (376;377). The elemental understanding of the nature of protein-protein interactions has begun to change, with novel research uncovering intrinsically unstructured regions of proteins forming important interfaces for a large number of protein-protein interactions (378;379). To drive comprehensive understanding of AGR2 function, protein-protein interaction studies have been undertaken to identify interaction partners; the logic being that most proteins are incorporated into complexes of other proteins to perform their specific function. Therefore, if the function of the complex or of any of the interacting partners is known, then inferences on the role of the studied protein can be made (292;380). A number of physiological interactors have been defined for AGR2, utilising yeast-2-hybrid, including metastasis associated C4.4A and α-dystroglycan (172), and the AAA+ chaperone and transcriptional regulator protein RuvBL2/Reptin. The latter of which has been validated both *in vitro* and *in vivo* (199). However, how these interactions might influence growth regulation is not clear. AGR2 can also interact with the extracellular receptor implicated in limb regeneration Prod1 (154), suggesting a link between extracellular protein trafficking or binding and cell migration. Thus, although there is no compelling mechanism yet established, that would define a key protein-protein interaction important for the oncogenic function of AGR2. Taken together these are four protein-interactors for AGR2 that are consistent with a role for AGR2 in promoting cancer associated phenotypes like cell growth and/or migration.

Two independent studies, concurrently with this study, have identified AGR2 as a homodimeric complex. Ryu *et al*. (202) identified an intermolecular disulphide bond *in vivo* in a human colon cancer cell line which could be chemically crosslinked through the single cysteine residue at position 81. Further, the dimerisation of AGR2 could attenuate ER stress induced cell death and only the dimeric complex could interact with the Hsp70 molecular chaperone and unfolded protein response signal pathway activator, BiP/GRP78. Agr2 mutant protein, containing cysteine-81 mutated to serine, exhibited reduced PERK and IREα induction, indicating that dimerisation is essential for the activation of the UPR signalling pathway. An attenuated interaction with BiP/GRP78, compared to the cysteine containing wild-type, was also evident. It is thought that, BiP/GRP78 is associated with PERK and IREα in the resting state and upon conditions of ER-stress, dissociates the complex triggering phosphorylation of IREα and PERK, resulting in downstream signal transduction (381;382). Thus, AGR2 appears to drive a form of allosteric regulation of the complex. As further validation of the intermolecular disulphide interaction, an *in vivo* cross-linking coupled to mass spectrometry approach has previously identified the AGR2 homodimer as the most abundant AGR2 complex (203).

In addition, a second possible dimerisation interface has been described during the course of studies presented herein. Nuclear magnetic resonance (NMR) studies aimed at defining the first illustration of the 3-dimensional organisation of AGR2, uncovered that mature AGR2 (lacking the N-terminal signalling sequence) spontaneously formed a monomer-dimer equilibrium (PDB code: 2LNS) (201). This heterogeneous nature was found to be shared with AGR3, a protein sharing 65% amino acid sequence identity. This study found that the cysteine-81 to serine mutant had minimal effect on dimerisation state, which is in contrast to previous *in vivo* studies (202). The monomer-dimer distribution model indicated a stable dimer with a dissociation constant of 8.83 μM. Using a truncated mutant to remove the unstable N-terminus of the protein, a dimerisation interface of $AGR2_{41\text{-}175}$ was identified through the anti-parallel arrangement of the α-1 (residues 58-67) helix (Figure 5-1). An alanine point mutagenesis strategy of residues comprising the arrangement, namely, E60, K63 and K64, identified E60A as a monomeric unit by NMR (PDB

code: 2LNT). The K64A-mutant also had disrupted dimerisation, as detected by analytical ultracentrifugation and gel permeation chromatography, confirming an E60-K64 salt bridge interaction essential in maintenance of dimer structure. This report also attributed the twenty amino acid unstructured N-terminal motif (21-40) as required for changes to the rates of cell adhesion, but not for influencing dimerisation. Only a limited number of PDIs exist as dimers, significantly, it should be noted that the parental ERP18 protein was originally thought to exist as a dimer through cysteine residues in the thioredoxin-domain (383); this was later disproven by NMR structural studies which determined ERP18 as a monomer (200). ERP18 shares the thioredoxin fold characteristic of AGR2 and AGR3 however lacks the tyrosine and lysine residues corresponding to positions 64 and 63 of AGR2 respectively, which may be enough to promote the monomeric structure.

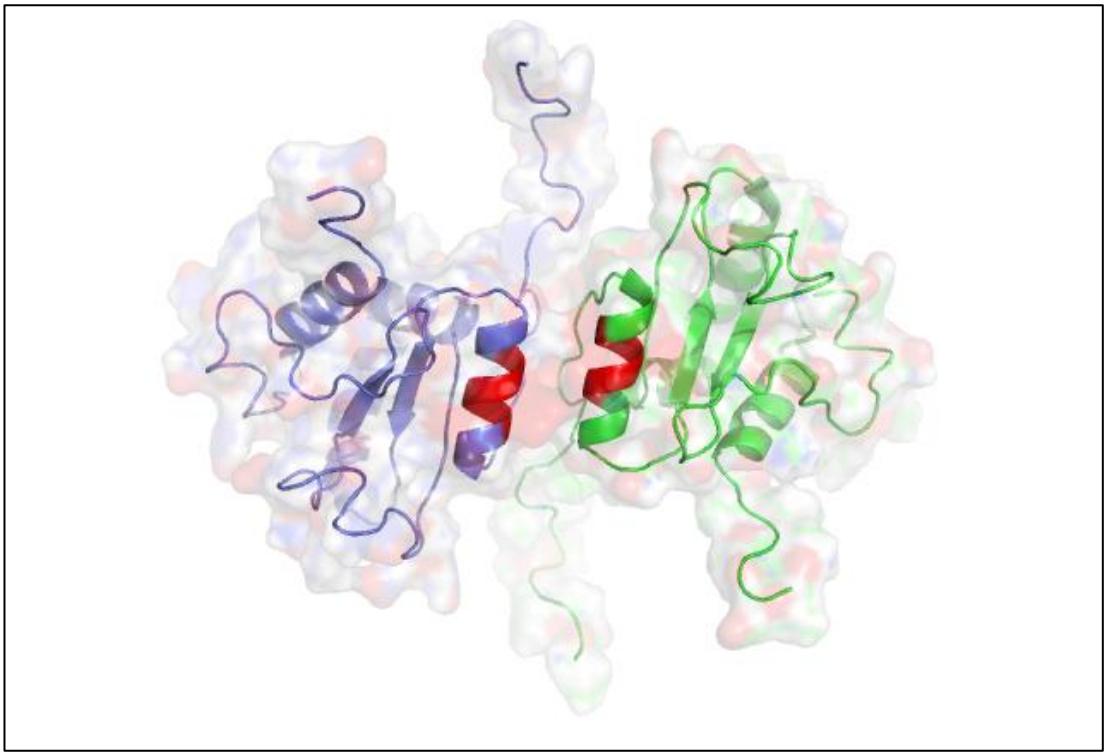


Figure 5-1 Illustration of AGR2 homodimer interface. Wild-type AGR2 dimer structure solved by NMR (PDB code: 2LNS). By point alanine mutation scan, the dimer interface is thought to form through an anti-parallel interaction between α-helices and stabilised through salt-bridge interactions of residues between the carboxylate group of E60 and the ammonium group of K64, with a hydrophobic interaction between Y63 residues having some influence on dimer stability. Highlighted are monomer 1 in blue and monomer 2 in green. The dimer interface defined by Patel *et al*. (201) of 60-EALYK-64 is indicated in red.

Structure-based lead molecule discovery remains a fundamental mechanism to exploit protein structure, develop small molecule leads and manipulate allostery in proteins (384). There are many examples of developing small molecule regulators of protein oligomerisation based on structure-function or *in silico* guided studies (385-388). Amongst the many emerging models in the cancer field, for example, small molecules exist that stabilise the tetramer-dimer/monomer conversion of pyruvate kinase that has the potential to regulate metabolism in cancer cells (389;390). This highlights the possibility to exploit the allosteric regulatory nature of many proteins through the development of small molecule tools or drug leads. The understanding of the quaternary structure of AGR2, how the monomer-dimer equilibrium influences its biomolecular activity and the development of tools to bias the equilibrium, guides an approach for reagents to interrogate AGR2 allostery and provide possible novel drug leads.

This chapter presents a His-tag based cloning and purification strategy allowing the purification of large quantities of recombinant, active AGR2. The protein was characterised and identified as a homodimer by size exclusion chromatography methods. During the scope of this research, structural data defining the 3-dimensional organisation of the AGR2 homodimer was published (201) coupled with a supplementary, somewhat conflicting report of AGR2 dimerisation (202), thus we set up a fluorescent microtiter based assay (two-site 'sandwich' microtiter assay, ^{2S}MTA) that measured AGR2 oligomerisation. This was used to determine whether the AGR2 dimer can be regulated *in trans* by self-peptides, and whether this information can be used to create mutants with altered dimer-monomer equilibrium. This information could then be used to evaluate the specific activity of such dimer mutants to determine whether the dimer is required for its protein-interaction functions. Further, the ^{2S}MTA could be further utilised for preliminary screens for natural compound molecules which could mediate the monomer-dimer equilibrium of AGR2 as novel drug leads for modulation of AGR2 function.

5.2 Results

5.2.1 Preparation of AGR2 pEHISTEV expression vector and purification of recombinant protein

Utilising a novel histidine-tagged expression vector methodology (pEHISTEV) developed by Liu and Naismith (286), an AGR2 expressing construct was cloned for protein purification (Figure 5-2A). The construct incorporates an amino-terminal hexa-histidine for nickel-based isolation from bacterial lysate, upstream of a tobacco etch virus (TEV) protease cleavage site. As a result the recombinant protein can be purified and the tag cleaved resulting in only a four amino acid residue overhang N-terminally of the desired protein sequence. The AGR2 sequence cloned into the expression plasmid lacked the disordered N-terminal cleavable signal peptide (Figure 1-4), such that the expressed product would be a representation of the mature functional protein from *in vivo* studies. Following transformation into *E. coli* BL21 (DE3) bacteria and selection, isolated colonies were cultured in LB and induced with IPTG. The induced his-tagged AGR2 fusion product, with a molecular mass of 21.2 kDa, was evident in crude bacterial lysate (Figure 5-2B). Following clearing of the bacterial lysate of membrane fragments and unlysed cells, the bacterial slurry was incubated with nickel-agarose beads. Beads were washed and eluted as described in Chapter 2.8.2 (Figure 5-2C). The eluted protein was pooled, dialysed to remove imidazole, and subject to a time course cleavage using TEV protease (Figure 5-2D). TEV protease incorporates a non-cleavable his-tag for downstream removal, and the fusion protease protein was present throughout as a 27 kDa band visualised by Coomassie staining. The fusion and cleaved AGR2 proteins, of 21.2 kDa and 18.3 kDa respectively, were also visible by Coomassie stain, and indicated that the His-tag was almost completely removed from AGR2 following one hour incubation. The cleaved protein solution was again run over a nickel column to remove the TEV protease, the cleaved tag and any uncleaved protein resulting in a preparation of pure cleaved AGR2 protein for further analysis.

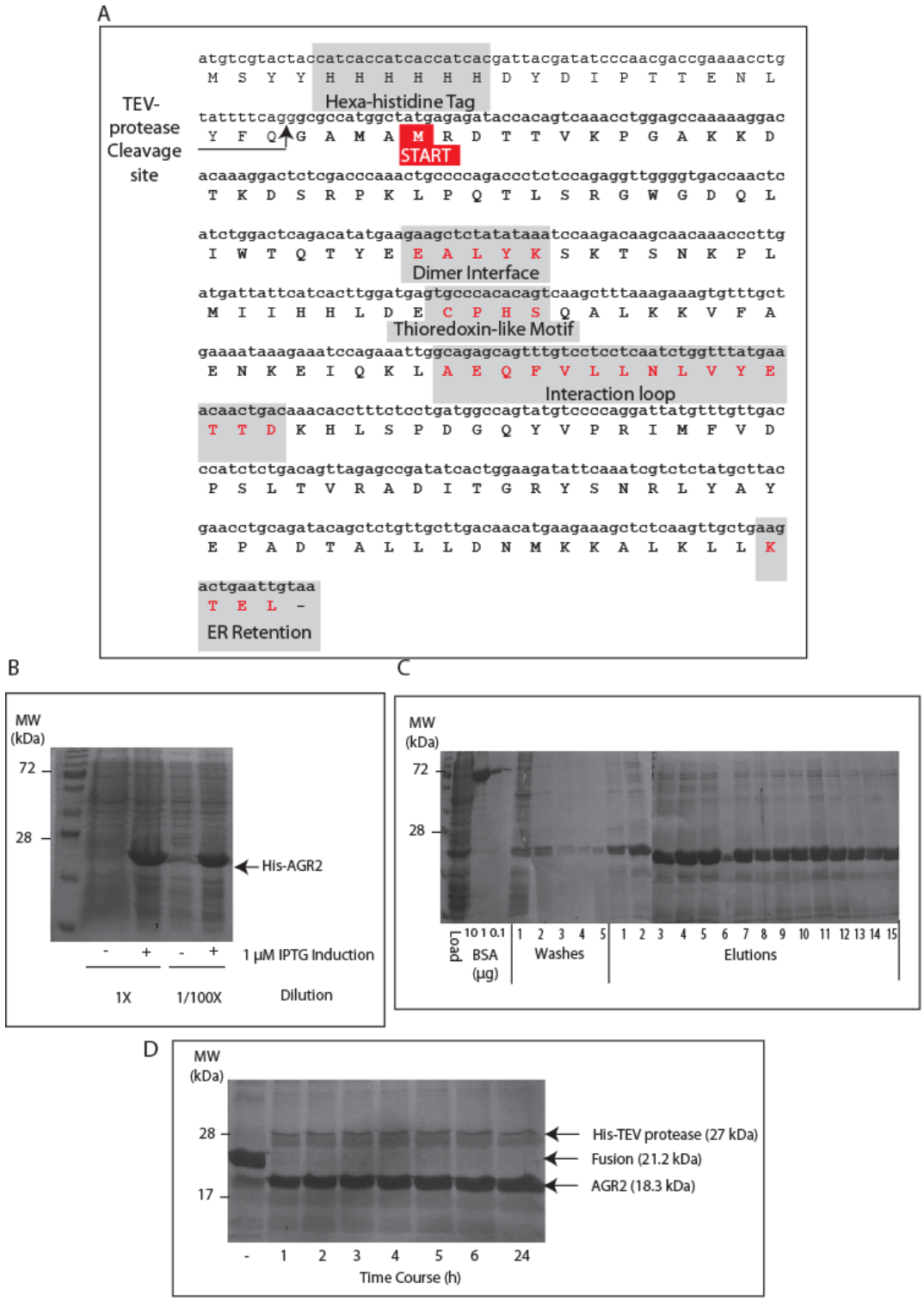


Figure 5-2 Purification strategy of recombinant AGR2. (A) Highlighted are the essential motifs of AGR2 including the E60-K64 dimerisation interface (201), the thioredoxin-like motif (202;215), the divergent substrate interaction loop (199) and the ER retention sequence (188). (B) SDS-PAGE separated IPTG-induced bacterial lysates expressing His-tagged AGR2 visualised with Coomassie

staining. A band representing the 21.2 kDa fusion protein is highlighted. (C) SDS-PAGE and Coomassie staining of wash and elution fractions of the His-AGR2 bacterial lysate incubated with nickel-agarose beads. To allow a visual estimate of protein concentration 10, 1 and 0.1 μg samples of BSA were loaded onto the gel also. (D) Purified protein was incubated with His-tagged TEV protease (10 μg per 1 mg protein, Life Technologies) and the cleavage carried out at 4^0C over a 24 hour time course. Following Coomassie staining, 3 distinct bands were evident the largest of 27 kDa represents the TEV protease, the 21.2 kDa corresponds to the size of the TEV-AGR2 fusion product, and the smallest 18.3 kDa band indicates the cleaved AGR2 product.

It was imperative following the purification protocol that that the recombinant protein was functional and that there was no detrimental effect on activity as a result of bacterial expression, the non-expression of the N-terminal signalling sequence, the additional 4 N-terminal amino acids and the cleavage of the fusion tag. Previously, it was published that AGR2 exhibited a defined specific affinity for interaction with a peptide aptamer containing the canonical TxIYY motif. This peptide could precipitate AGR2 from crude lysate, and as a result could reactivate p53 activity *in vivo* (187;275). The 12mer (HLPTTIYYGPPG) and truncated 6mer (PTTIYY) have previously been identified as exhibiting the optimal affinity for AGR2 binding, therefore these were utilised in an ELISA based assay to confirm the activity of the purified recombinant AGR2 protein. Biotinylated peptide aptamers were immobilised on the substrate of a streptavidin-coated 96 well plate and incubated with the titrated recombinant protein, prior to detection with an AGR2 specific antibody (Figure 5-3). Compared to control conditions (no peptide and non-interacting peptide), AGR2 exhibited an increase of specific binding to the interacting peptides of between 5-6 fold. Exemplifying that recombinant AGR2 purified in this method retains a previously published biochemical activity (187;275)

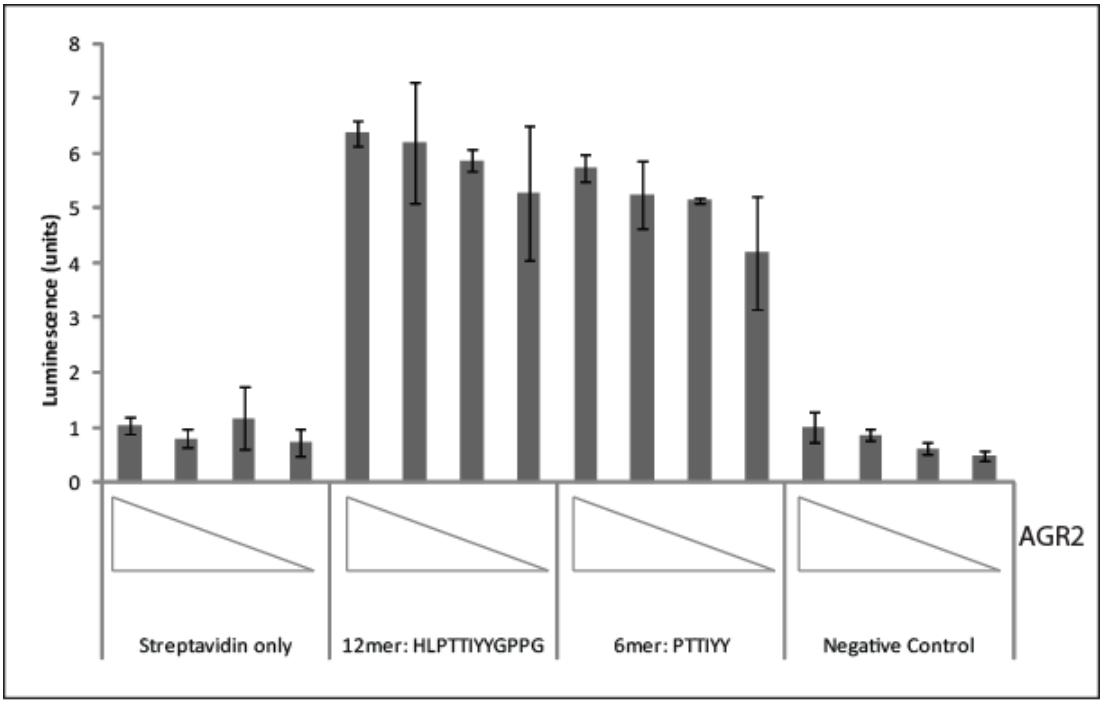
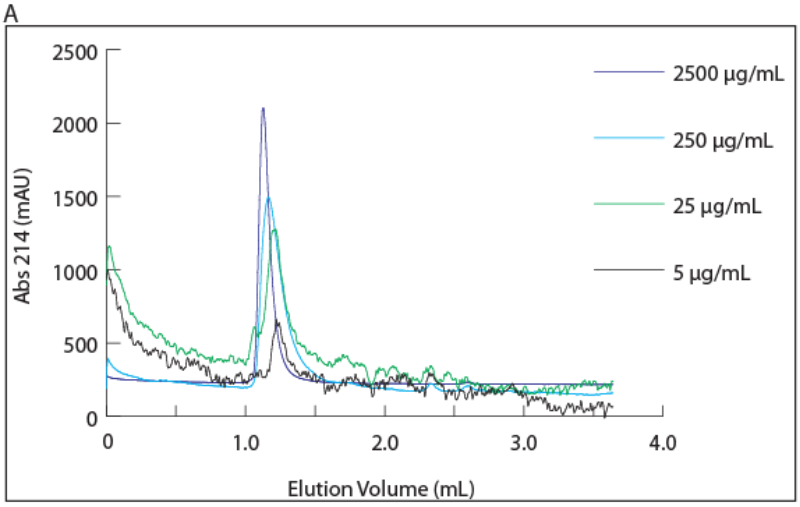


Figure 5-3 ELISA based analysis of recombinant AGR2 interaction with biotinylated peptide aptamers. 96-well plate was coated with streptavidin, prior to incubation with 5 µg per well of biotinylated peptide. Following incubation, washing and blocking steps, a dilution series of AGR2 protein was applied (1000, 500, 250 and 125 ng) and incubated. Interaction was detected with primary AGR2 polyclonal antibody (K47) and swine anti-rabbit HRP conjugated secondary antibody.

Gel-filtration size exclusion chromatography of purified recombinant mature $AGR2_{21-175}$ protein was used to provide insight into the quaternary structure of the protein (Figure 5-4A). At highest concentrations (2500 μg/mL, 136 μM), the protein eluted with a calculated molecular mass of 34.8 kDa (Figures 5-4B), significantly greater than the theoretical monomeric molecular mass of 18.3 kDa (Figure 5-4C, Expasy Compute pI/MW tool, http://web.expasy.org/compute_pi/), more closely resembling a dimer structure relative to the theoretical mass of 36.5 kDa (Figure 5-4C). The relationship between the concentration of protein injected and integrated area of the GF peak is linear, with an R^2 of 0.9995 (Figure 5-4D). The elution profiles of globular proteins used for calibration are summarised in Figure 5-4E and 5-4F. By comparison, a prior analysis of $AGR2_{21-175}$ by SEC-MALLS gave a dimeric mass of 30.5 kDa from a protein with a monomeric mass if 17.8 kDa and a predicted dimeric mass of 32.2 kDa (201). AGR2 was diluted from 2500 μg/mL (136 μM) to 250 μg/mL (13.6 μM), 25 μg/mL (1.36 μM) and 5 μg/mL (0.27 μM) prior to injection on the Sephadex-75 column to determine whether there is a concentration-dependence to dimerisation (Figure 5-4A, 5-4B ad 5-4C). AGR2 protein injected at a concentration of 250 μg/mL eluted with an estimated mass of 32.4 kDa, while protein at 25 μg/mL exhibited a slower eluting species with an estimated mass of 29.1 kDa and 5 μg/mL protein injected resulted in a calculated molecular mass of 26.1 kDa, suggesting that the protein can exist in a dimer-monomer equilibrium as it approaches the predicted monomeric mass of 18.3 kDa at lower concentrations. These data are consistent with reports suggesting that a stable dimeric species of AGR2 exists (201;202) and requires an *in vitro* concentration of the pure protein in the low μM range (201).

A



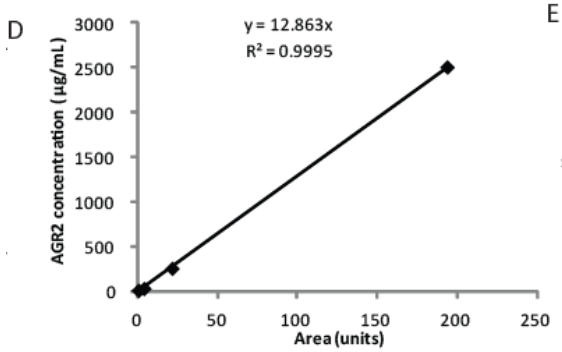
B

Concentration (µg/mL)	Ve (mL)	Kav	Calculated log MW	Predicted MW (Da)
2500	1.13	0.201258	4.54	34843
250	1.15	0.213836	4.51	32429
25	1.18	0.242704	4.46	29119
5	1.21	0.251572	4.42	26147

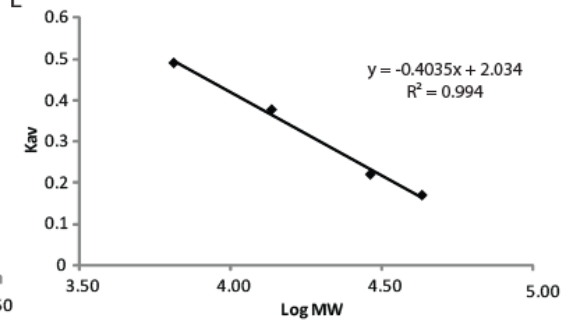
C

	MW (Da)	log MW	Calculated Kav	Predicted Ve (mL)
Monomer	18252.1	4.26	0.31456	1.31
Dimer	36504.2	4.56	0.19309	1.12

D



E



F

Protein	MW (Da)	Log MW	Ve (mL)	Kav
Ovalbumin	43000	4.63	1.08	0.169811
Carbonic Anhydrase	29000	4.46	1.16	0.220126
Ribonuclease A	13700	4.14	1.41	0.377358
Aprotinin	6500	3.81	1.59	0.490566

Figure 5-4 Investigating the quaternary structure of AGR2$_{21\text{-}175}$. (A) Gel filtration size exclusion chromatography trace of AGR2$_{21\text{-}175}$ protein, injected at the indicated concentrations, using an analytical Superdex PC 75 and detected by ultraviolet absorbance at wavelength of 214 nm (corresponding to the peptide bond). Traces have been normalised for ease of visual comparison. (B) Calculation of molecular mass of the peaks of AGR2 using the elution volume (C) The calculation of the elution volume of the monomeric (18.3 kDa) and dimeric (36.5 kDa) species of the recombinant protein. (D) Illustration of the integrated area of the AGR2 protein peak (from A), as a function of AGR2 protein concentration at the time of injection, highlighting the highly significant linearity between protein absorbance upon elution (at 214 nm) and protein (concentration) injected. (E & F)

Calibration of the Superdex PC 75 column with the indicated gel filtration markers (GE) in SEC buffer (50 mM Tris pH 7.5, 200 mM NaCl). These could be plotted to allow the calculation of the linear relationship between the logarithm of the molecular weight and the partition coefficient (Kav, where Kav = elution volume/void volume, void volume of Superdex PC 75 is 0.81 mL).

5.2.2 Characterisation of antibody reagents for AGR2 detection

Due to the significant similarity between AGR2 and AGR3 proteins, a panel of monoclonal antibodies were raised against AGR3 with the aim of defining the level of independence of expression between AGR2 and AGR3 in tumour biopsies (248). These four antibodies were subject to an ELISA-based study to determine their specificity for cross-reactivity between AGR3 and AGR2 (Figure 5-5). Two of the antibodies raised to be specific to AGR3, MAB3.1 and MAB3.2 (Figures 5-5A and 5-5B), indeed showed selectivity for the AGR3 protein, as titration of the antibody resulted in increased detection when AGR3 was immobilised on the substrate. Comparatively, these did not have any increased affinity above background for AGR2. However, through these studies it was apparent that one monoclonal antibody, MAB3.4, exhibited selective cross-reactivity with AGR2 (Figure 5-5C and 5-5D). This cross-reactivity, while not beneficial for *in vivo* studies for the differentiation of AGR2 from AGR3, could be utilised for *in vitro* studies where only the recombinant protein of interest is present.

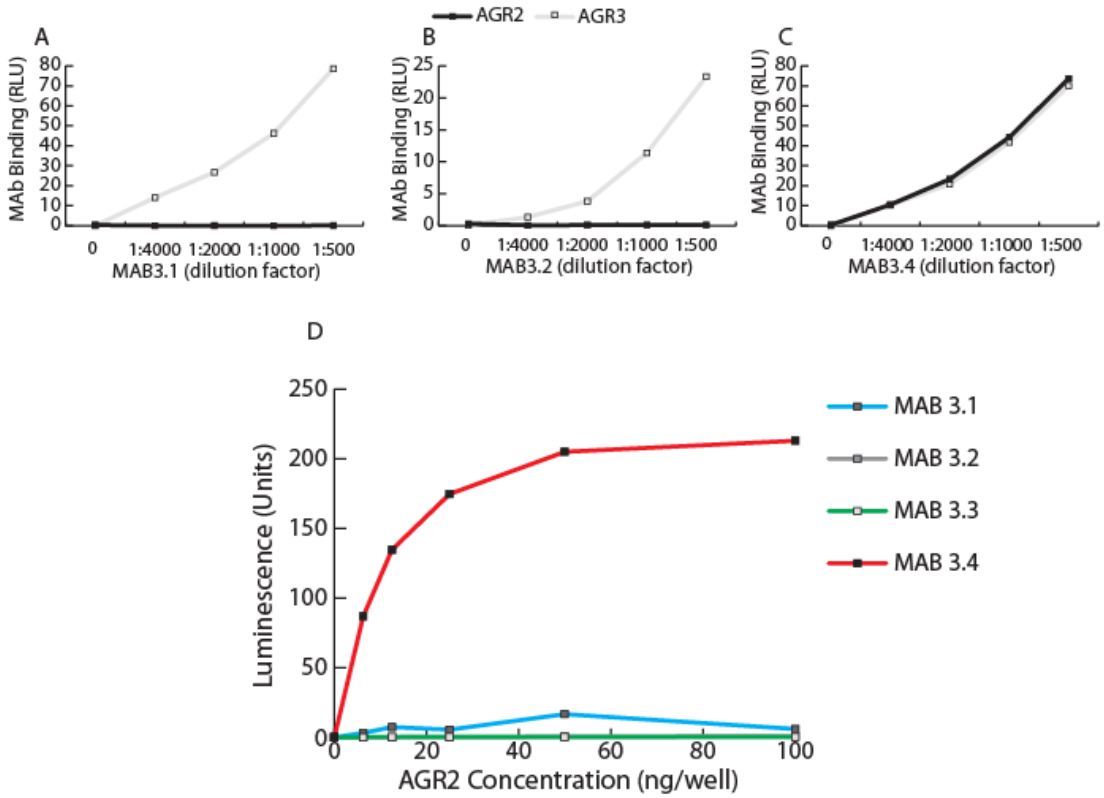


Figure 5-5 Analysis of bespoke monoclonal antibodies raised against AGR3 for cross-reactivity with AGR2. (A and B) MABs 3.1 and 3.2 AGR3 specific monoclonal antibodies bind to AGR3 protein (black line) immobilised on to the ELISA plate substrate but not the AGR2 (grey line). (C) One of the monoclonal antibodies, MAB 3.4, exhibits cross-reactivity with AGR2. (D) The cross reactivity of MAB3.4 (red line) was confirmed by immobilising the monoclonal antibodies on the substrate and titrating recombinant AGR2.

It was then endeavoured to fine map the AGR2 epitope recognised by MAB3.4 using an overlapping peptide scan. 15mer biotinylated peptides from the AGR2 (and AGR3) primary sequence, with a 5 amino acid overlap, were synthesised incorporating an SGSG spacer (Table 5-1). The peptide library was immobilised in a 96 well ELISA plate and incubated with the MAB3.4 antibody (Figure 5-6A). Non-specific interactions were removed with multiple wash steps prior to incubation with an HRP-conjugated rabbit anti-mouse secondary antibody. Concurrently, the alternative commercial anti-AGR2 monoclonal antibody (Abcam) was also epitope mapped (Figure 5-6B). However, this antibody exhibited a discontinuous non-linear epitope, thus the recognition site on AGR2 could not be identified. The MAB3.4 antibody showed a robust binding to a linear motif present in peptides 6 and 7 from AGR2 and peptides 21 and 22 of AGR3. The shared sequence present in the 4 positive peptide hits was PLMII, from AGR2, and PLMVI, from AGR3 (Figure 5-6C) resulting in the mapping of the MAB3.4 epitope as PLMxI (Figure 5-6D) (independently, phage display data confirmed these epitopes, data not shown, (248)). The epitope for the MAB3.4 antibody was also found to be spatially distinct from the proposed AGR2 dimerisation interfaces (201;202) (Figure 5-6E). By thoroughly characterising the AGR2 epitope recognised by MAB3.4, we could identify that the epitope (residues 71-75) was distinct to the perceived dimerisation interfaces of residues 60-64 (201) and cysteine-81 (202). It was therefore considered whether this nature could be utilised in the development of assays to measure AGR2 oligomerisation status.

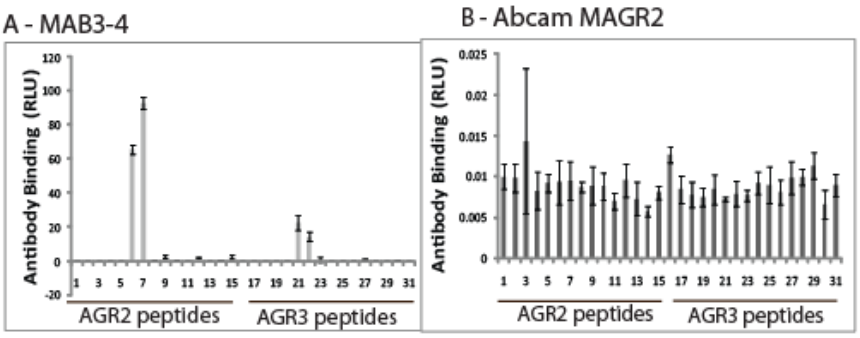
Currently, it is not clear which of the functions of AGR2 the oligomeric state might regulate. It is described that monomeric AGR2 lacks the ability to interact with BiP/GRP78 (202) activating the UPR, yet this is reliant on the formation of the dimer through cysteine-81. However, given that an additional dimer interface (201) has been described, this may impinge on the conclusions of the aforementioned study. In order to develop sensitive and quantitative assays to measure AGR2 protein specific activity in a variety of protein-protein interactions, we sought to develop an assay that measures, in a microplate format, the oligomeric structure of AGR2. Such a small-scale assay could facilitate determining whether more information could be

derived from the competing dimerisation interface hypotheses, while also investigating if interacting proteins or additional allosteric effectors can alter the oligomerisation of the protein.

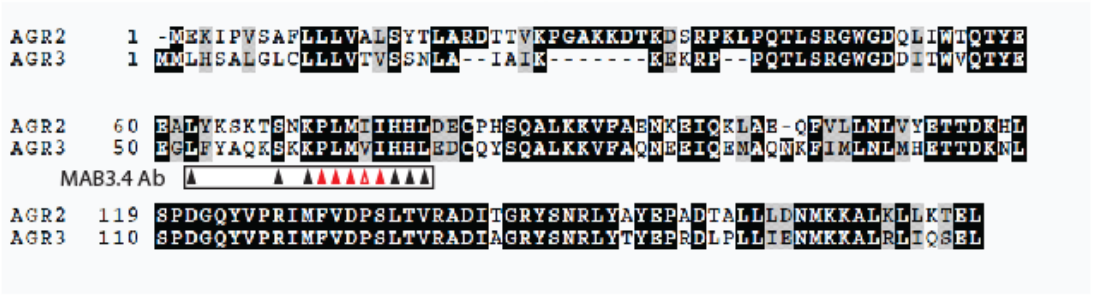
Table 5-1 Overlapping peptide libraries of AGR2 and AGR3 consisting of biotinylated 15 amino acid residue peptides, with 5 amino acid overlap, from the primary sequence of AGR2 (1-16) and AGR3 (17-31). The C-terminal residue of the putative cleavage site of the N-terminal signalling sequence has been highlighted in red.

Peptide Number	AGR2 Sequence	Residue Number
1	LLVALSYTLARDTTV	11-25
2	RDTTVKPGAKKDTKD	21-35
3	KDTKDSRPKLPQTLS	31-45
4	PQTLSRGWGDQLIWT	41-55
5	QLIWTQTYEEALYKS	51-65
6	ALYKSKTSNKPLMII	61-75
7	PLMIIHHLDECPHSQ	71-85
8	CPHSQALKKVFAENK	81-95
9	FAENKEIQKLAEQFV	91-105
10	AEQFVLLNLVYETTD	101-115
11	YETTDKHLSPDGQYV	111-125
12	DGQYVPRIMFVDPSL	121-135
13	VDPSLTVRADITGRY	131-145
14	ITGRYSNRLYAYEPA	141-155
15	AYEPADTALLLDNMK	151-165
16	LDNMKKALKLLKTEL	161-175

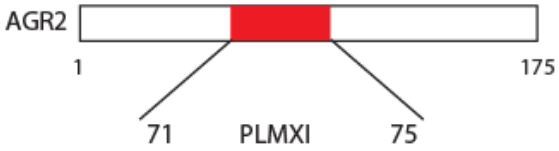
Peptide Number	AGR3 Sequence	Residue Number
17	LLVTVSSNLAIAIKK	12-26
18	IAIKKEKRPPQTLSR	22-36
19	QTLSRGWGDDITWVQ	32-46
20	ITWVQTYEEGLFYAQ	42-56
21	LFYAQKSKKPLMVIH	52-66
22	LMVIHHLEDCQYSQA	62-76
23	QYSQALKKVFAQNEE	72-86
24	AQNEEIQEMAQNKFI	82-96
25	QNKFIMLNLMHETTD	92-106
26	HETTDKNLSPDGQYV	102-116
27	DGQYVPRIMFVDPSL	112-126
28	VDPSLTVRADIAGRY	122-136
29	IAGRYSNRLYTYEPR	132-146
30	TYEPRDLPLLIENMK	142-156
31	IENMKKALRLIQSEL	152-166



C



D



E

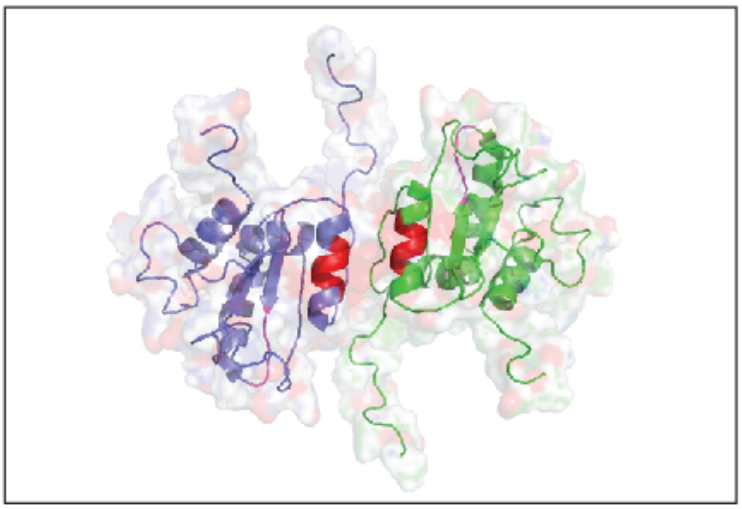


Figure 5-6 Epitope mapping of monoclonal anti-AGR2 antibodies. (A) Peptide scan (Table 5-1) ELISA of AGR2 and AGR3 highlighting the linear motifs recognised by the MAB3.4 antibody.

Peptides 6 and 7 or AGR2 and peptides 21 and 22 of AGR3 were identified as binding significantly above background. (B) The peptide scan of the commercial monoclonal anti-AGR2 antibody (Abcam) did not identify a linear epitope, rather this antibody recognises a discontinuous epitope conformed by higher order protein structure. (C) Box shade alignment (http://www.ch.embnet.org/software/BOX_form.html)of AGR2 and AGR3 proteins highlighting the shared residues of AGR2 and AGR3 from peptides 6, 7, 21 and 22 (black arrows), the red arrows indicate which of these are also shared among the 4 peptides. (D) Illustration of the derived AGR2 epitope recognised by MAB3.4 as 71-PLMxI-75. (E) Illustration of the spatial distribution of the 60-EALYK-64 dimerisation interface (201) (red) and the MAB3.4 epitope (248) (pink) PDB code: 2LNS.

5.2.3 Developing a quantitative microtiter assay to measure AGR2 oligomerisation

We aimed to determine whether a quantitative two-site sandwich microtiter assay (^{2S}MTA) could be used to quantify oligomerisation. Such an assay would then be amenable to evaluate, in the first instance, whether self-peptides of the interfaces could abolish the multimeric structure through direct competition, whether any of these self peptides may function *in trans* to allosterically regulate the structure of AGR2 and, utilising chemical libraries for small molecules and/or natural products, could alter the stability of the oligomer. The premise of the ^{2S}MTA is that the same immobilised monoclonal antibody used to capture an AGR2 molecule could then be used to detect the corresponding epitope in the liquid phase if the protein was oligomeric, e.g. monomers cannot be detected by this assay (Figure 5-7). Using a fluorescent labelling method would allow for quantitative detection of oligomers over monomers. This capture assay is limited in distinguishing monomeric arrangement from oligomeric arrangement. Therefore, it can only comment on the formation of higher order structures and not definitively ascribe a dimer.

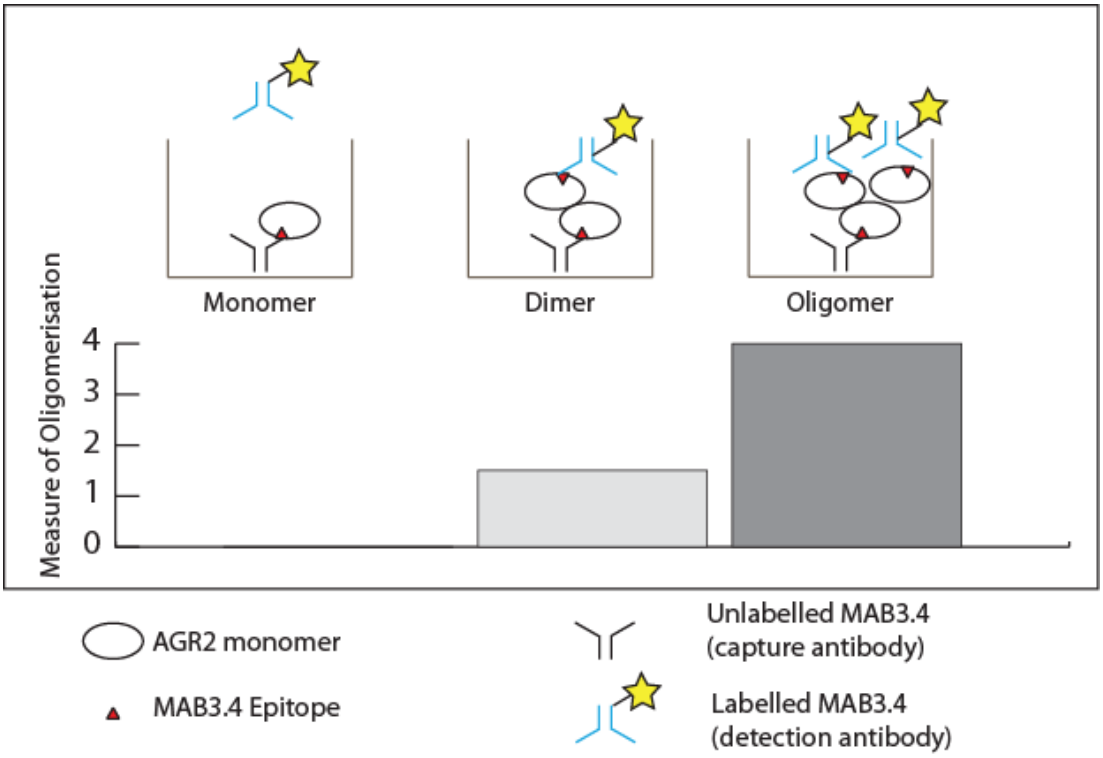


Figure 5-7 Theoretical illustration of the ELISA-based oligomerisation assay. The predicted generic emission of an ^{2S}MTA to measure AGR2 oligomerisation using MAB3.4 to both immobilise the AGR2 protein on the solid phase, and, through a fluorescent labelling step, also detect the epitope on the corresponding interaction partner(s).

In order to determine whether this monoclonal antibody could be used in an assay to measure AGR2 oligomerisation, a near infrared DyLight800 (DyL800, Thermo) fluorophore conjugation strategy, together with the Licor Odyssey imaging system for detection, was used to label a population of the MAB3.4 antibody. The labelling reaction proceeds through NHS (N-hydroxysuccinimide) ester of the fluorophore reacts randomly through modification of primary amines of the lysine residues side chain, forming a stable covalent amide bond between the antibody and the fluorophore. This enables capture of AGR2 using immobilised unlabelled MAB3.4, and detection of oligomeric (dimeric or larger order) complexes using the DyL800-labelled version of the same MAB3.4 antibody. Consequently, this allows quantitation of the extent of oligomerisation, as the detection of higher order assemblies is proportional to the amount of DyL800-MAB3.4 bound (as per Figure 5-7).

To develop this ^{2S}MTA, the fluorescent probe labelling strategy must not interfere with the interaction of MAB3.4 with recombinant AGR2 protein. It is essential that the amide bond between lysine and DyL800 probe does not diminish the antibodies complementary determining region (epitope recognition site) such that labelled MAB3.4 retains functionality. A titration of MAB3.4 and DyL800-MAB3.4 revealed that the labelled antibody retained affinity for its epitope in AGR2 protein and that binding could be quantified in a dose dependent manner (Figure 5-8A and 5-8B). The AGR2 protein-monoclonal antibody interaction in Figure 5-8B measured the binding of the MAB3.4 or DyL800-MAB3.4 using peroxidase-labelled anti-mouse IgG secondary antibody and detection using enhanced chemiluminescence. The DyL800-labelled antibody appeared to have a slight attenuation of affinity compared to the unlabelled control, which could be anticipated since the antibody has been covalently modified. Importantly, however, it retained considerable selectivity for AGR2.

Following this, we evaluated the binding reaction using Licor Odyssey instrument and near infrared detection by excitation at 750 nm and fluorophore emission at 800 nm (Figure 5-8C, Figure 5-9). The binding of DyL800-MAB3.4 to AGR2 protein adsorbed onto the solid phase could be detected in a positive correlation manner with

increasing AGR2 concentration (Figure 5-8D), relative to the unlabelled MAB3.4 which did not give rise to any signal. The binding reaction appeared more linear using DyL800-MAB3.4 in the direct excitation-emission assay (Figure 5-8D) than when the indirect peroxidase-labelled secondary antibody assay (Figure 5-8B) was used. There are two explanations for this phenomenon. Firstly, the chemiluminescence assay (in Figure 5-8D) uses "enhanced" enzymatic conversion by the antibody-peroxidase conjugate of substrate to create light that is quantified as a function of a fixed time. As a result, the data do not reflect a reaction rate but a final product accumulation. This "enhancement" of the primary signal may produce results that deviate from linearity. By contrast, the DyL800-conjugated monoclonal antibody emits a signal that is detected directly and the results are presumably more linear. Secondly, the indirect enhanced chemiluminescence assay incorporates an additional 75 minutes of incubation with secondary antibody, washings, and chemiluminescence substrate addition and this incorporates an unquantified effect of antibody off rate on the signal intensity. By contrast, after incubations with the DyL800-conjugated antibody, the reaction is washed rapidly, and fluorescence intensity is read immediately thus minimising effects of the antibody off rate on the signal intensity. This validates the DyL800-MAB3.4 excitation emission assay as a bioactive probe.

As such, we next evaluated whether DyL800-MAB3.4 could detect oligomeric forms of AGR2 in a sandwich-ELISA style quantitative assay to measure oligomerisation of the protein using an unlabelled population of MAB3.4 as the capture antibody (as outlined in Figure 5-7). A summary of the approach using DyL800-MAB3.4 to detect an oligomeric (potentially dimeric) species of AGR2 is presented in Figure 5-8E. Full reactions were assembled with or without the adsorption of unlabelled MAB3.4 on the solid phase to compare the relationship between MAB3.4 capture and DyL800 detection. Subsequent incubation with titrated recombinant AGR2 protein allowed immobilisation of the protein by the capture antibody, followed by detection with fixed concentration of DyL800-MAB3.4 with near IR fluorescent quantitation. A dose-dependent signal could be observed (Figure 5-8F), suggesting that this assay can indeed capture and measure levels of oligomeric AGR2 protein.

It can be concluded from these experiments that: (i) The DyL800-labelling does not abrogate the AGR2 epitope recognition by the MAB3.4 monoclonal antibody. (ii) The DyL800-label appropriately labels MAB3.4 such that the complex linearly detects AGR2 protein, and this interaction can be detected using near-IR methods. (iii) Finally, that MAB3.4 adsorbed onto the substrate can immobilise AGR2 from the solution phase, and detection of higher order structures can be quantified by the interaction of DyL800-MAB3.4 with the epitope recognition site of oligomeric subunits.

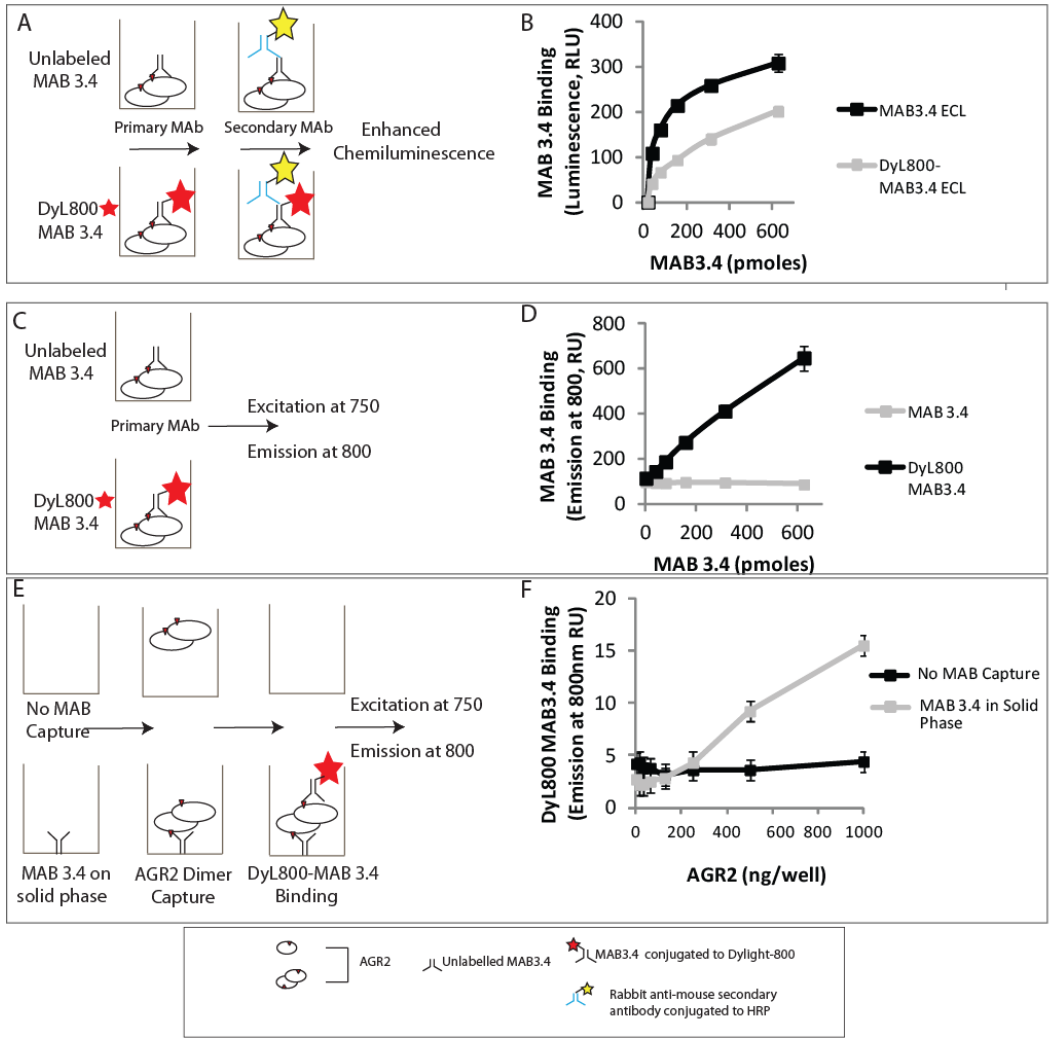


Figure 5-8 Developing a quantitative microtiter assay to measure AGR2 oligomerisation. (A) Evaluation of the bioactivity of fluorescently labelled MAB3.4 in a luminescence based ELISA. (B) Titration of unlabelled MAB3.4 and DyLight800-conjugated MAB3.4 in the mobile phase into wells of a 96-well plate adsorbed with AGR2 in the solid phase. The binding of MAB3.4 to AGR2 was measured using an anti-mouse IgG secondary antibody conjugated to horseradish peroxidase. The bioactivity of the monoclonal antibody (in RLU) is measured as a function of increasing MAB3.4 concentration. (C) Assessing of the bioactivity of fluorescently labelled MAB3.4 in a near infrared detection assay. (D) Unlabelled and DyL800-labelled MAB3.4 were titrated into wells containing AGR2 adsorbed in the solid phase. The monoclonal antibody-AGR2 complex was detected and measured as levels of emission at 800 nm (following excitation at 750 nm) as a function of antibody concentration. (E) The ability of unlabelled MAB3.4, in the solid phase, to capture mobile recombinant AGR2. (F) Unlabelled MAB3.4 was coated onto the 96-well plate, and increasing amount of AGR2 protein titrated to allow capture onto the solid phase. Fixed amounts of DyL800-MAB3.4 were added into reactions to detect the epitope of interacting units. The extent of

oligomerisation was quantified as levels of emission at 800 nm as a function of increasing AGR2 protein concentration. All data is presented as a mean of triplicate titrations (± standard deviation) and is representative of multiple labelled DyL800-MAB3.4 populations.

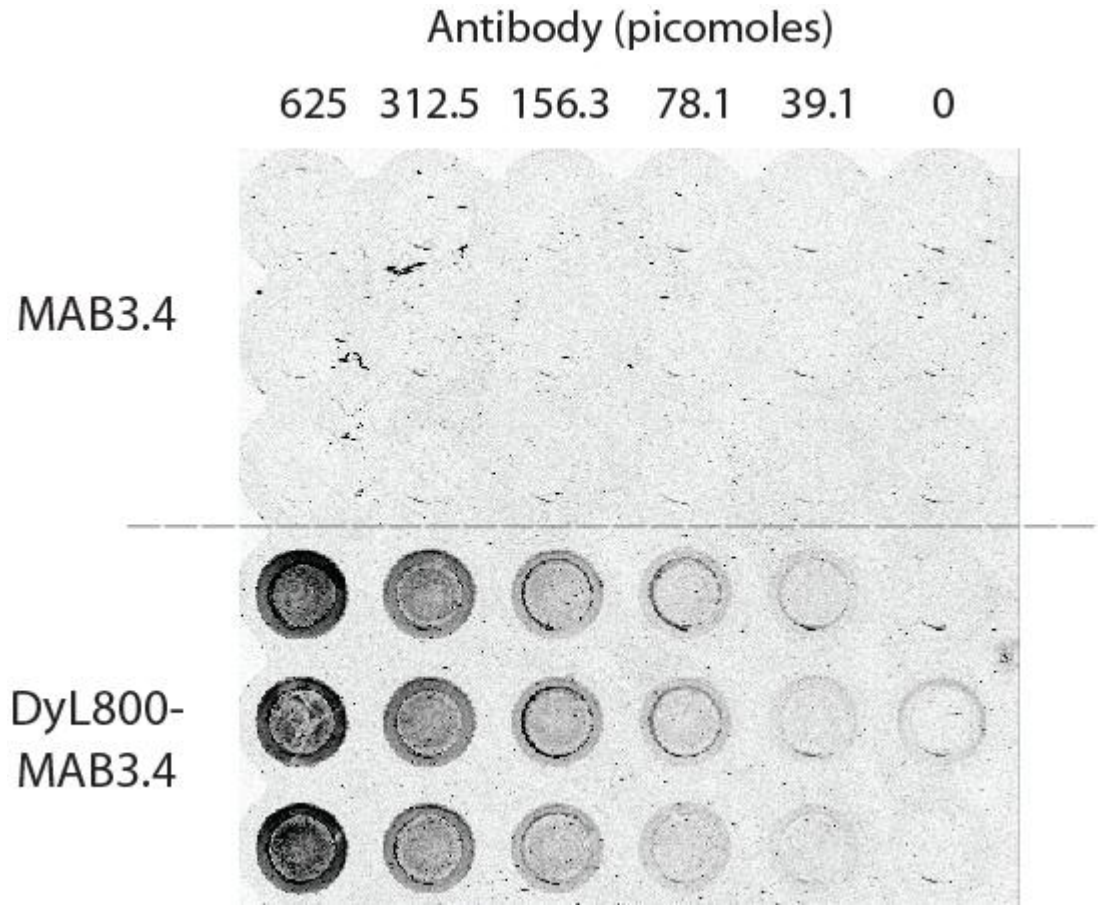


Figure 5-9 Confirmation of DyL800-labelling of MAB3.4 and specificity for AGR2. Visualisation of the level of AGR2 detection in the solid phase by titrated unlabelled and DyL800-labelled MAB3.4 and detected by excitation at 750 nm and emission at 800 nm, converted to black and white for clarity (As per Figure 5-8C and 5-D).

5.2.4 Identification of linear motifs which stabilise or disrupt the oligomer equilibrium of AGR2.

The literature describes that AGR2 can exist as a dimer (201-203) (Figure 5-4), thus antibody reagents and an assay to quantify AGR2 oligomerisation have been developed. Subsequently, it was sought to determine whether the oligomeric equilibrium of AGR2 can be attenuated using dimer interface self-peptides or indeed stabilised *in trans* by regulatory motifs (Figure 5-10). In stage I of the assay, oligomeric AGR2 is captured on the surface of the well using the immobilisation properties of MAB3.4 (Figure 5-8E and 5-8F). Utilising the overlapping peptide library derived from the open reading frame of AGR2 (Table 5-1) in stage II, self peptides are pre-incubated with the recombinant protein allowing interaction with the oligomeric complex. Finally, in stage III the DyL800-MAB3.4 fluorophore conjugated to the antibody could be used to detect the exposed epitope of the interacting subunits allowing quantitation of the AGR2 population which is in the oligomeric state. Thus, allowing the identification of the effect of interaction of AGR2 proteins with linear motifs which might allosterically contribute to protein structure and dimerisation could be identified.

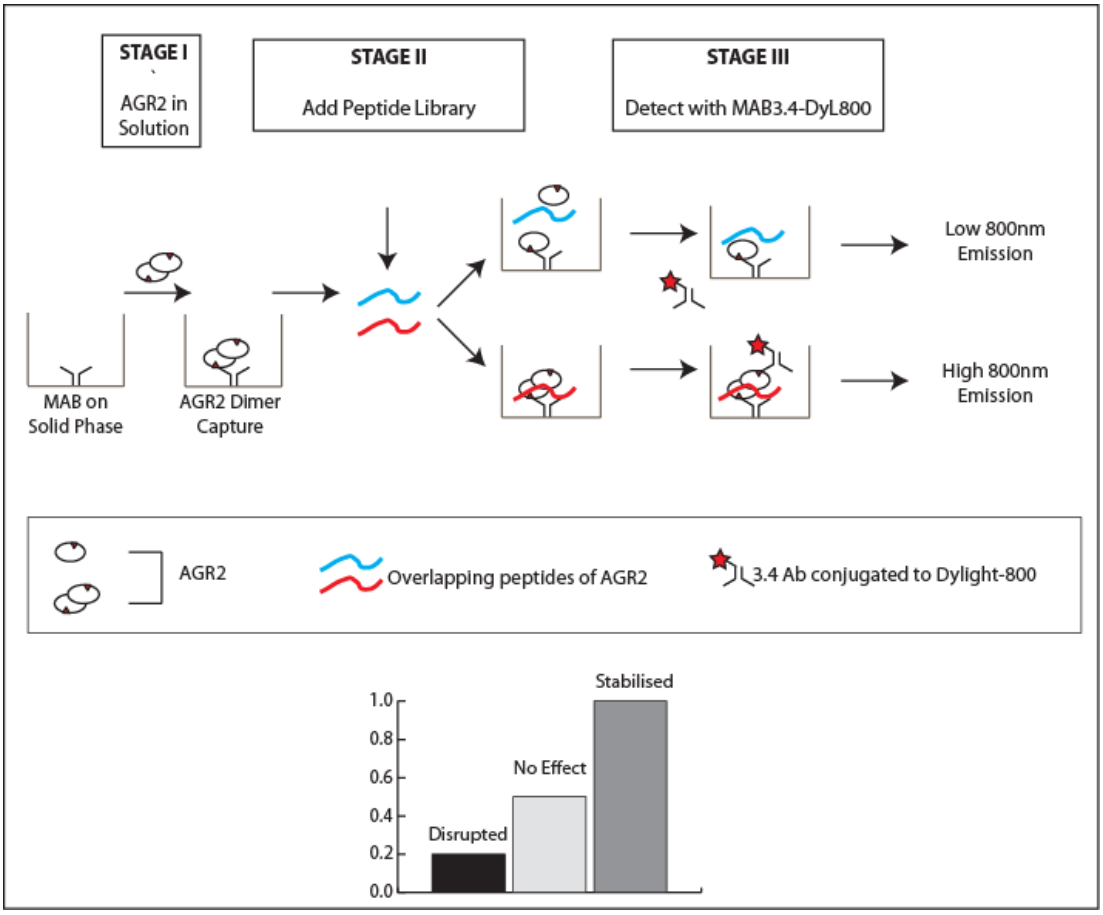


Figure 5-10 Illustration of the hypothesis of using the ^{2S}MTA coupled with an overlapping peptide library of the AGR2 protein to identify peptides which function *in trans* to allosterically regulate the proteins oligomeric state. The immobilisation of the recombinant protein with the same antibody used to detect it (MAB3.4) means that the level of fluorescent dye detected is a function of the level of AGR2 protein quaternary structure.

Indeed, the incubation of AGR2 (5.5 pmol) with self-peptides (1.25 nmol) at significant molar excess resulted in perceived allosteric modulation resulting in variation in the level of oligomer detected (Figure 5-11). Relative to the oligomerisation state of the DMSO control, several peptides indicated significant stabilisation of the oligomer, while others indicated slight effects in dissociation of the dimer (Figure 5-11A). Incubation with peptides 6 and 7 (Figure 5-6) increased binding of the labelled DyL800-MAB3.4. This was surprising as these peptides contain the epitope site for recognition by MAB3.4 antibody and were therefore expected to compete with the linear binding site of the recombinant AGR2 protein for interaction with MAB3.4. If this was the case, a reduction in capture by unlabelled MAB3.4 on the plate and/or detection by DyL800-MAB3.4 would have been observed. However this was not the case. It was hypothesised that the increase in detection was not definitively as a result of stabilisation of the oligomer, rather might be an effect of modification of the epitope recognition for the antibody. Thus peptide 6 and 7 were not subject to further investigation in the study. Stabilisation of the multimer was evident in conditions where the protein has been incubated with peptides 1, 2 and 3 of AGR2, indicating that the linear motifs of these peptides may function as allosteric promoters of the oligomer. These peptides are derived from the N-terminal disordered region of AGR2 (Figure 1-4A) which provides an interesting insight into the effect of disordered motifs in the allosteric modulation of the protein. As 11 of 15 residues of peptide 1 are cleaved *in vivo* during the trafficking of the protein to the endoplasmic reticulum, this peptide was also removed from future deeper analysis. Titration of peptides 2 and 3 (Figure 5-11B) indicated a dose dependency of oligomerisation stabilisation. This was, particularly evident in the case of peptide 3, at the concentration ranging from 50 to 12.5 mM. High levels of variation in lower concentrations of peptide contribute to loss of sensitivity in detection. Peptides 2 and 3 incorporate the majority of the unstructured N-terminal domain identified in Figure 1-5, and thus because of recent research of the level of protein disorder in allosteric regulation and the utilisation of this in drug discovery leads (391-393), highlighted it as a potential regulatory motif essential for the stabilisation of the quaternary structure.

The remaining peptides do not appear to have a significant influence on the oligomeric equilibrium of AGR2, with only minor modulations detected particularly in the C-terminal portion of the protein. However, it is important not to discard these without further consideration. Peptides overlapping the dimerisation interface do not stabilise the signal in the assay. Rather, peptides 4 and 5 reduce the fluorescence (Figure 5-11A) with peptide 5 containing the interface sequence 60-EALYK-64 inducing the greatest degree of oligomer de-stabilisation. In fact, further titration of the peptides surrounding dimer interface (Figure 5-11C) indicate peptide 5 as a significant destabilising factor in the oligomeric state of AGR2. The ability of peptide 4 to similarly destabilise the oligomer suggests other motifs in AGR2 in close proximity can be targeted to affect the intersubunit E60-K64 interface. The common amino acid sequences of peptides 4 and 5 that attenuate the oligomer stability include 51-QLIWT-55 (Figure 5-12A), whilst the more bioactive peptide has the additional 56-QTYEEALYK-64 linear motif that comprises the dimer interface. This may explain why peptide 5 is more active at disrupting the AGR2 oligomer using the ^{2S}MTA.

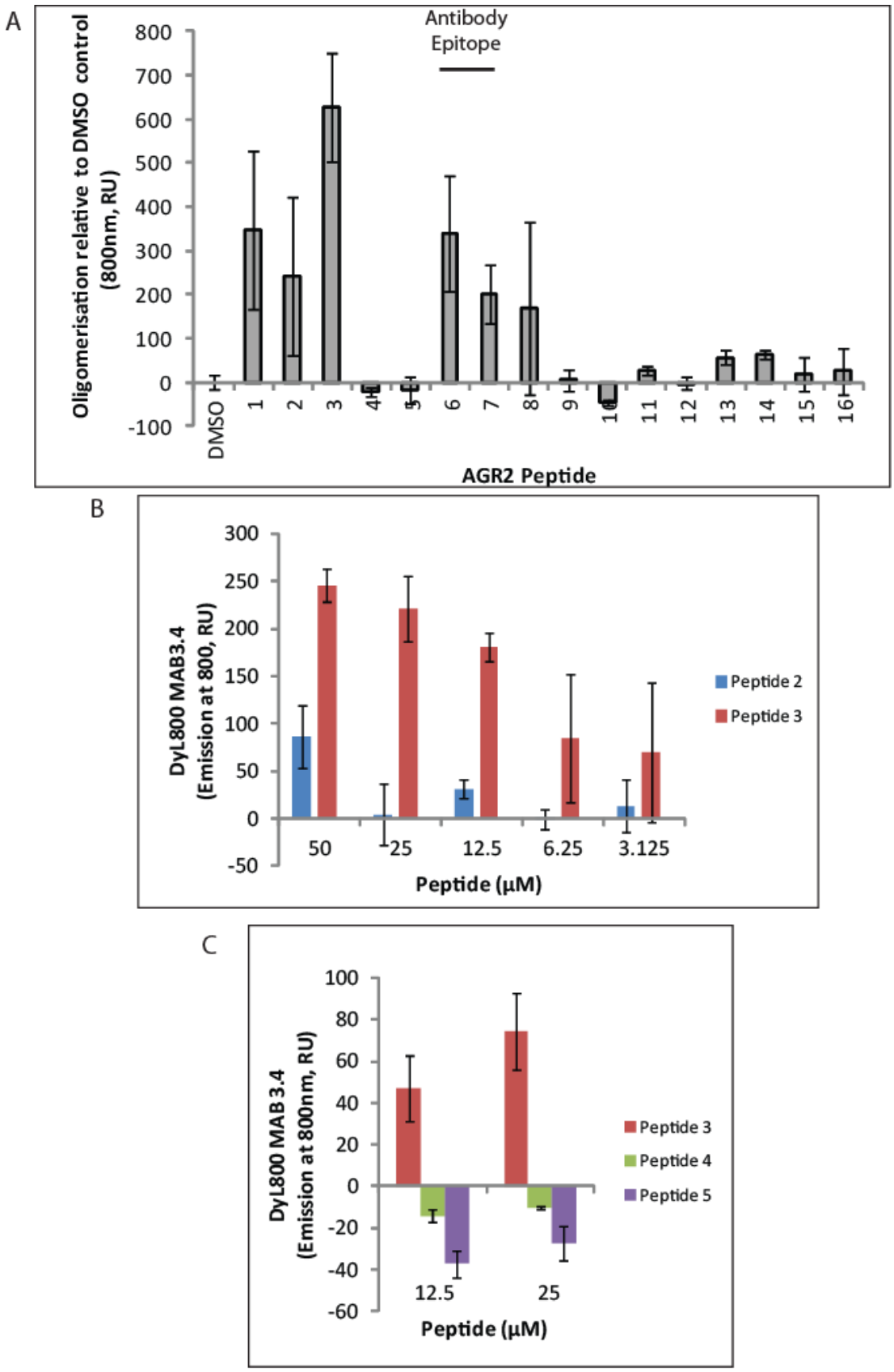

A
Antibody Epitope
Oligomerisation relative to DMSO control (800nm, RU)
800
700
600
500
400
300
200
100
0
-100
DMSO
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
AGR2 Peptide
B
DyL800 MAB3.4 (Emission at 800, RU)
300
250
200
150
100
50
0
-50
50
25
12.5
6.25
3.125
Peptide (µM)
Peptide 2
Peptide 3
C
DyL800 MAB 3.4 (Emission at 800nm, RU)
100
80
60
40
20
0
-20
-40
-60
12.5
25
Peptide (µM)
Peptide 3
Peptide 4
Peptide 5

Figure 5-11 Quantification of the effect of self peptides on the oligomerisation state of AGR2 (A) Peptides derived from the open reading frame of AGR2 (Table 5-1) were incubated at a mass of 1.25 nmol with 5.5 pmol recombinant AGR2 protein prior to being subject to ^{2S}MTA oligomerisation assay. Briefly, after washing, the DyL800-MAB3.4 probe was added for 1 hour and the extent of oligomerisation plotted as a function of fluorophore emission at 800 nm. Data is presented normalised to the no peptide (DMSO) control and the mean of three replicates with ± standard deviation. Highlighted are the peptides which derive the MAB3.4 antibody recognition site. (B) A titration of peptides 2 and 3 from the N-terminal disordered region indicates a dose dependent stabilisation of the oligomeric AGR2 population. Peptides were titrated from 50-3.125 μM (5-3.125 nmoles per well) normalised to DMSO control ± standard deviation. (C) Titration of peptides 3 (stimulatory), 4 (inhibitory) and 5 (inhibitory, incorporating the 60-EALYK-64 motif. Indicating that increased concentration of peptides 4 and 5 reduce the detection of the oligomeric species.

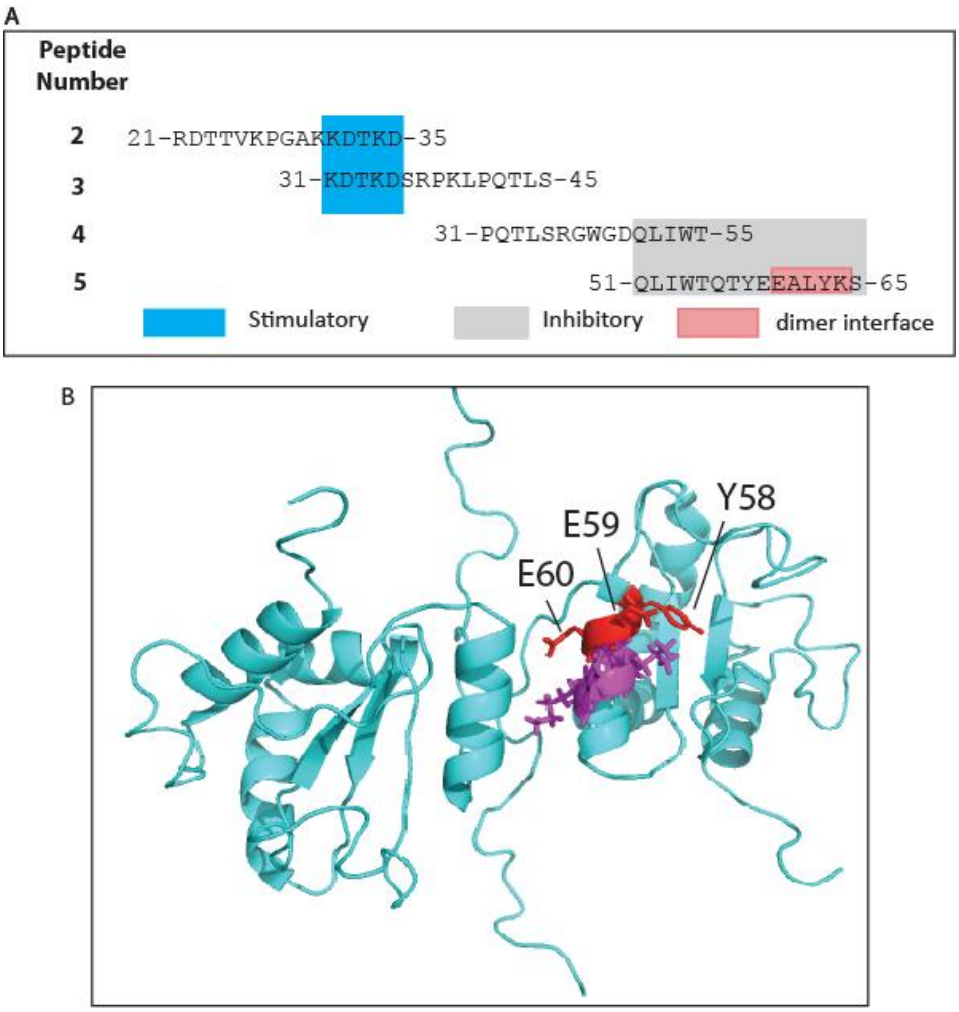


Figure 5-12 Defining and visualising the residues which comprise the AGR2 homodimeric interface. (A) Illustration of the sequence of bioactive peptides (from figure 5-11) derived from the disordered and dimerisation regions of AGR2 protein. Highlighted are minimal motifs predicted to drive stimulation (KDTKD from peptides 2 and 3) and inhibition (QLIWT from peptides 4 and 5). The dimer interface (201) in peptide 5 60-EALYK-64 is shaded pink. (B) Summary of the positions of residues Y58, E59 and E60 across the dimer interface (PDB: 2LNS). The whole EALYK motif is highlighted in purple, but the key residues (bulky hydrophobic or charged) in close proximity for mutation coloured red

Utilising the knowledge of motifs from self-peptides which could regulate the oligomerisation of AGR2, a panel of mutants was created to analyse the application of the ^{2S}MTA on full length mutant proteins to further study changes in oligomerisation state. Visualising the AGR2 structure using NMR structural data (Figure 5-12B) (PBD: 2LNS) (201), we chose to mutate bulky hydrophobic, and highly charged amino acid residues from the interface inferred by self-peptide disruption. Tyrosine residue 58 was mutated to alanine, since it protrudes into the adjacent internal β-sheet and might be required for stabilising the dimer interface. In addition, residue E59 was mutated as it protrudes outward into the solvent, while E60 forms an intramolecular salt bridge with K64 across the adjacent monomer and constitutes an important feature of the dimer interface (201). Also created was an N-terminal truncation mutant lacking the initial 45 amino acid residues (Δ45), as this served two purposes: both peptides 2 and 3, covering residues 21-45, had a stabilising effect on the AGR2 oligomer (Figure 5-11), and the N-terminal region up to residue 45 has been identified as unstructured by the disorder predictor software (Figure 1-5) as showing a high level of disorder. Therefore, removal of these residues removes the potential regulatory domain which may stabilise the protein. These mutants would allow progression of oligomerisation modulation, identified through self-peptide analysis, and subsequent identification and validation of residues essential for the formation and stabilisation of the dimer interface in whole proteins.

The mutant DNA constructs were created by site directed mutagenesis (Y58A, E59A and E60A), while the truncated mutant (Δ45) required re-cloning into the pEHISTEV backbone. These were then transformed into *E.coli* strain BL21 (DE3) and successfully induced with IPTG, as previously (Figure 5-13A). Bacteria were then lysed, cleared and 10 µL of each sample was separated on a 12% polyacrylamide gel, before transferred to nitrocellulose. Immunoblots were then probed using DyL800-MAB3.4, and detected on the 800-channel of the Licor Odyssey, allowing relative quantitation of the AGR2-isoform concentration from the crude bacterial lysate using Odyssey Sa software package (Licor) (Figure 5-13B). The crude lysates were then normalised to the same concentration of induced protein for downstream

experiments. This crude lysate normalisation allowed rapid and efficient testing of mutant AGR2 constructs compared to complete nickel-affinity chromatography purification of each isoform.

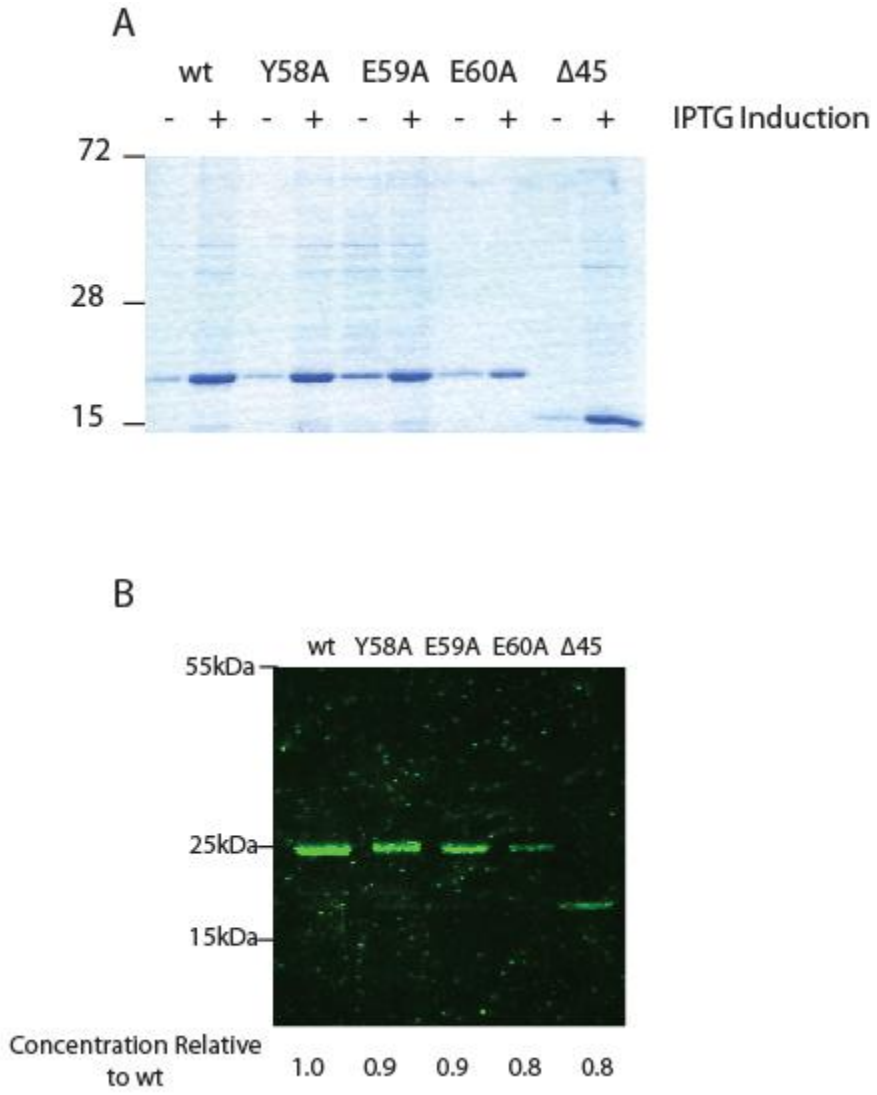


Figure 5-13 Creation of AGR2 mutants to investigate the modulation of oligomer stability. (A) Coomassie Blue stain of bacterial samples of wild-type AGR2, the three alanine point mutants, Y58, E59 and E60, and the N-terminal truncation, Δ45, pre- and post- IPTG induction, showing the induction of expression of the protein isoforms of interest. (B) Following lysis and clearing, 10 µL of each crude lysate was separated by SDS-PAGE, transferred to nitrocellulose and immunoblotted using DyL800-MAB3.4 and direct detection by excitation at 750 nm and emission at 800 nm. Densitometry of bands using Odyssey Sa (Licor) allowed quantitation of protein concentration and normalisation as indicated.

A titration of AGR2-E59A revealed it has an oligomer stability profile similar to that of wt-AGR2 (Figure 5-14A), which might be expected since the side chain of the E59 is thought to project away from the interaction interface toward the solvent (Figure 5-12B). Thus, mutation to alanine is having minimal effect on the stability of the quaternary structure. In the wild type protein, the Y58 side chain projects into the secondary structure motif of the interaction partner for possible stability. The data revealed that the AGR2-Y58A mutant was attenuated as an oligomer with a signal in the ^{2S}MTA only seen at higher concentrations of the protein (Figure 5-14A). This indicates that Y58 may not be an essential residue for interaction to form the homodimer, but that it may perform an additional stabilising constraint assisting the dimer formation. When comparing the AGR2-E60A mutant protein oligomeric state with that of the wt-AGR2, demonstrates a clear diminishing of the detection of the oligomer. The mutation of glutamate-60 to alanine abrogates the described salt-bridge formed between E60 of one monomer with K64 of the second subunit (201). This glutamic acid residue is thus thought to be essential for the stability of the oligomeric complex. By contrast to these loss-of-function mutants AGR2-Δ45 exists as a more stable oligomer. The enhanced stability of the oligomer by deletion of the disordered region is consistent with prior research showing that deletion of the amino acids 21-40 increased dimer stability and facilitated structural analysis of AGR2 as a dimeric species (201).

The presentation of oligomerisation data of wt-AGR2, AGR2-E60A and AGR2-Δ45 as a function of increasing protein concentration (Figure 5-14B) demonstrates the contrasting effect of the removal of essential residues in the formation of the higher order assemblies. The mutation of E60 clearly has a significant detrimental effect on the formation of oligomers, whereas deletion of the unstructured N-terminal residues allows realisation of an oligomer with increased stability compared to the quaternary structure of the wild-type protein.

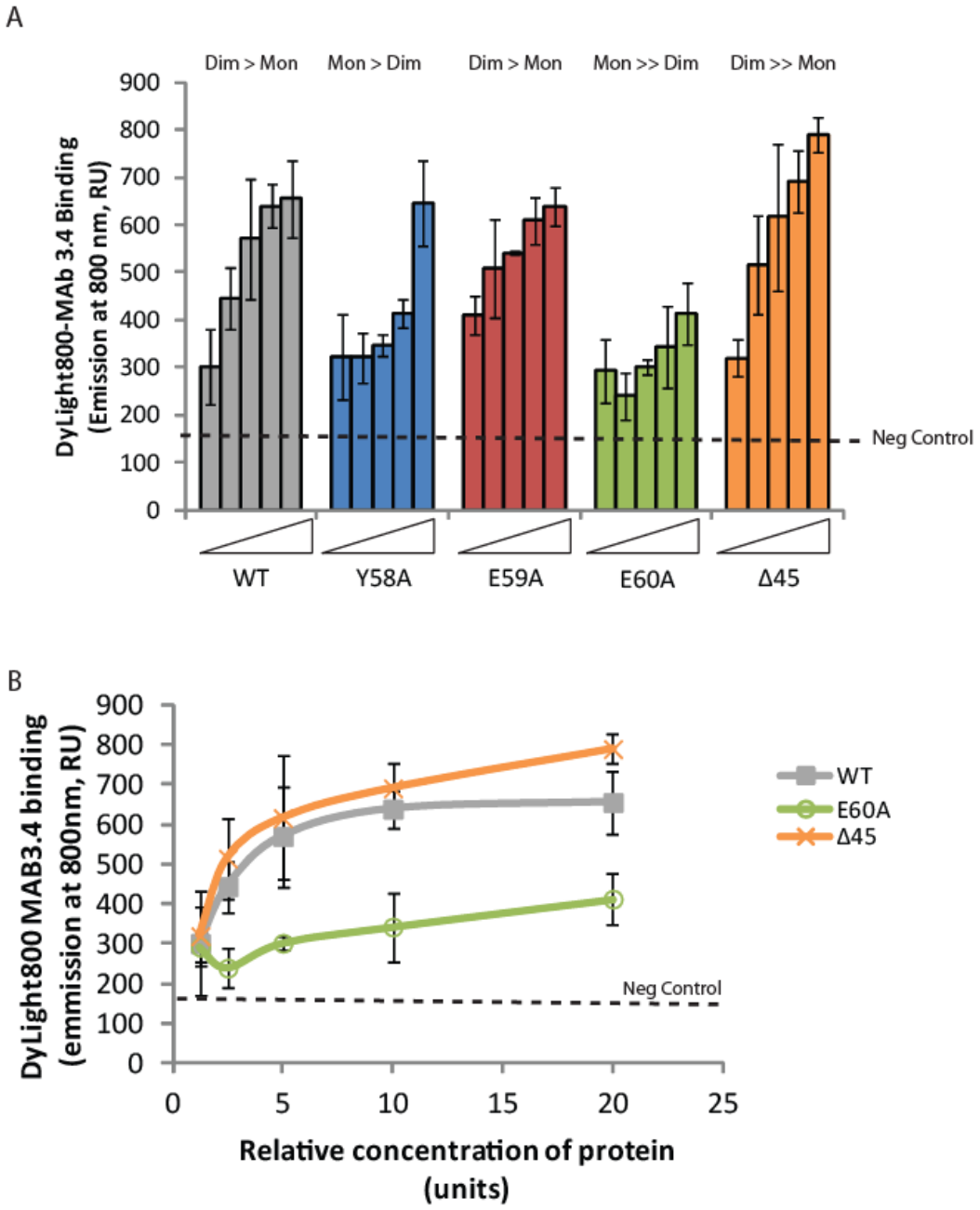


Figure 5-14 Determination of motifs influencing dimer stability of AGR2 protein using a mutagenesis and truncation strategy. (A) Dimer stability derived by ^{2S} MTA in the AGR2 mutant panel (wild-type, Y58A, E59A, E60A and Δ45). Increasing amounts of normalised wild-type and mutant AGR2 induced crude bacterial lysate (20-, 10-, 5-, 2.5- and 1-fold normalised) were captured using unlabelled MAB3.4 immobilised on the plate. Following one hour incubation, the extent of oligomerisation was quantified using DyL800-MAB3.4 coupled to emission of the fluorophore at 800 nm. The data is presented as an average of triplicate titrations, and representative of two repeated experiments. The no protein control background signal is represented by the dashed line. The ratio of

perceived dimer to monomer is described above. (B) Oligomerisation data for wt-, E60A- and Δ45- are presented as a function of protein concentration indicating the evidence for essential (E60) and destabilising (N-terminal) motifs of AGR2 affecting the quaternary structure.

A secondary assay was sought to validate the conclusions of linear motifs influencing AGR2 oligomerisation derived from the ^{2S}MTA. The use of a DSS chemical crosslinking assay has previously been described to covalently trap dimeric AGR2 species and detected using Western blotting (202;203). Mapping of this using mass spectrometry has identified lysine residue 95 as the dominant crosslinking site (203). This knowledge was used to determine whether the dimeric mutant panel exhibited altered dimerisation using the DSS crosslink assay (Figure 5-15). A titration of DSS into reactions containing constant normalised concentration of wt-AGR2, AGR2-E60A and AGR2-Δ45 confirmed that AGR2-E60A is attenuated in dimer formation in response to DSS compared to the wild-type protein (Figure 5-15B compared to 5-15A). At all DSS concentrations, the AGR2-E60A mutant was discovered to be less readily covalently linked to form a dimer (and other higher order species) than the wild-type protein, indicating that the K95 residues were not within the 11.4 Å proximity of the DSS linker as frequently. This suggests that the dimer was less available for crosslinking and consequently less stable during the incubation. In addition, the Δ45 N-terminal truncated mutant exists in a more stable, spontaneous dimer, even in the absence of crosslinker under these conditions (Figure 5-15C lane 1). This implies that deletion of the N-terminus can allow the formation of the stable dimer even in the presence of 0.1% SDS of the running gel. It was subsequently discovered that the titration of DMSO into recombinant AGR2-Δ45 drove the stabilisation of the dimer at 10% DMSO (Figure 15-5D). The effect of DMSO on protein stability has been observed before (394), however it is intriguing to discover that the Δ45 mutant can be significantly stabilised by high concentrations of DMSO such that it is not disrupted in a denaturing gel separation system. This may give an insight into the importance of the disordered N-terminal linear motif in the allosteric regulation of AGR2 protein. This is consistent with the results of the ^{2S}MTA as well as with the recent publication of AGR2 structure (201), showing that the deletion of the N-terminal 40 amino acids of the protein stabilises the dimer (201). These data together demonstrate that Δ45-AGR2 forms a more stable dimer whilst AGR2-E60A is predominantly monomeric.

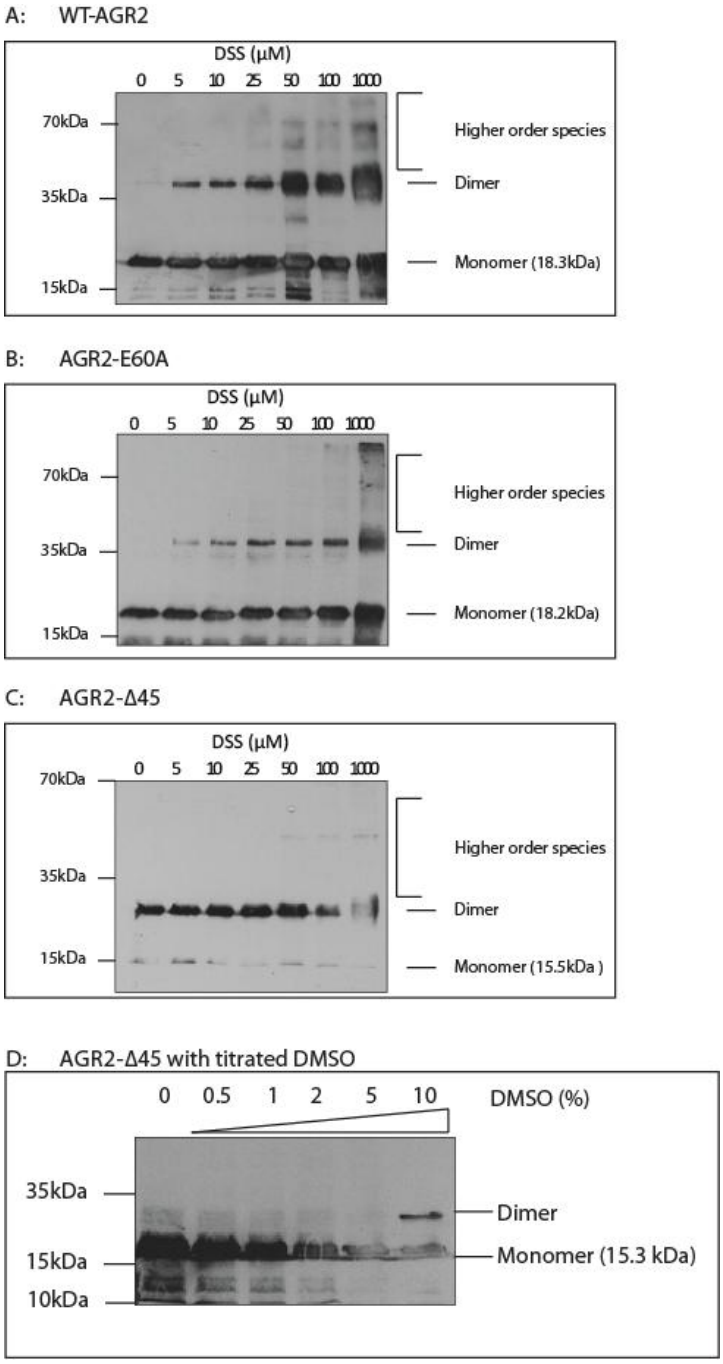


Figure 5-15 The effect of DSS on the dimerisation of AGR2 mutants. The indicated proteins (A) wt-AGR2, (B) AGR2-E60A and (C) AGR2-Δ45, were incubated with increasing concentrations of DSS and separated by electrophoresis on a 0.1% SDS gel. AGR2 was detected using MAB3.4 by immunoblotting. The positions of monomeric, dimeric and higher order species are highlighted. (D) AGR2-Δ45 was incubated with increasing concentrations of DMSO prior to immunoblot as above, indicating a stabilising effect of the dimeric species at high concentrations of DMSO.

5.2.5 Evaluating the role of the AGR2 dimer

The only described function of the AGR2 homo-dimer is the recent publication of the dimeric species being essential for the interaction with BiP/GRP78, allowing the AGR2-dependent mediation of the ER stress signalling pathway (202). However, the monomeric mutant was dependent in that study on the mutation of the cysteine-81 residue removing the intermolecular disulphide interaction and did not take into account the 60-EALYK-64 subsequently reported (201). Having created a series of AGR2 mutant proteins with altered oligomeric equilibrium, it could be investigated whether these have any gain-of-function or loss-of-function role in known AGR2 activities. The activities investigated were the specific binding of wt-AGR2 to the TxIYY peptide consensus docking motif (187;275) and the interaction with the AAA+ chaperone protein Reptin (199). A titration of wt-AGR2, AGR2-E60A and AGR2-Δ45 indicates that all three exhibited relatively similar activity in binding to the peptide substrate (Figure 5-16A). This implies that, with respect to the consensus peptide binding site on AGR2 (currently undefined at the structural level), peptide binding is independent of quaternary structure. By contrast, the Reptin-AGR2 interaction is substantially stabilised using the gain-of-function, more stable dimer mutant AGR2-Δ45, whilst the monomer-biased AGR2-E60A mutant is attenuated in Reptin binding, when compared to the wt-AGR2-Reptin relationship (Figure 5-16B). The importance of the AGR2-Reptin interaction is not fully elucidated at present, however study of their co-expression in tumour biopsies is underway and AGR2-Reptin expression is linked in commonly used cancer cell lines (199), indicating a possible role of AGR2 allosteric modulation in cancer progression studies.

These data provide the biochemical evidence to suggest that AGR2 protein dimerisation can play a role in its affinity for a client protein. Since there are several other binding proteins identified as AGR2 interactors by yeast-2-hybrid (169), this panel of AGR2 oligomerisation mutants can be used to access the role of the AGR2 dimer in protein-protein interactions *in vitro* and in cells.

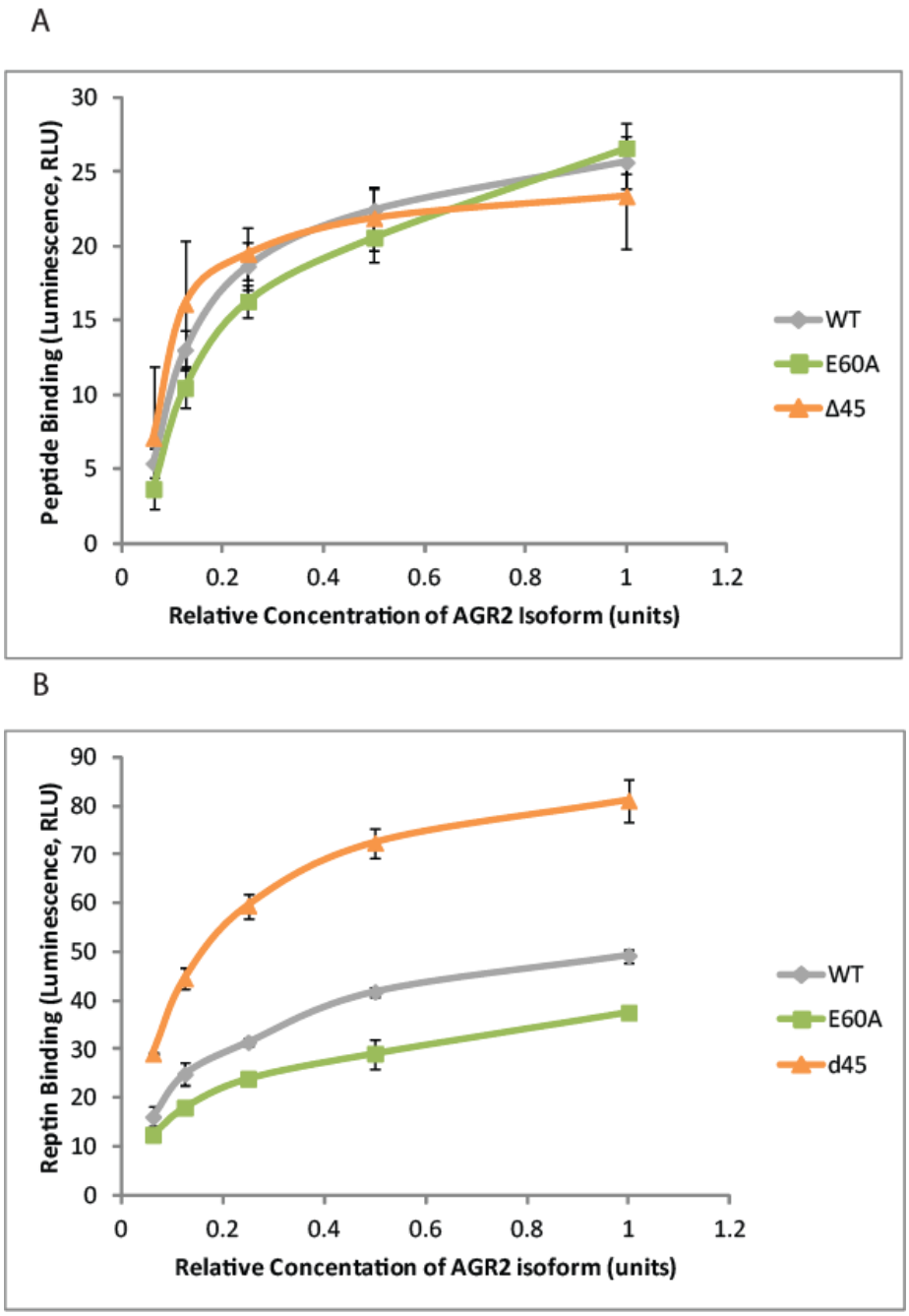


Figure 5-16 Analysis of the biochemical function of AGR2 mutants with diverged oligomer-monomer bias. Examination of the interaction of wild-type AGR2 and oligomerisation/dimerisation mutants with (A) the biotinylated TxIYY consensus binding motif and (B) the AAA+ chaperone protein Reptin. TxIYY or Reptin was adsorbed in the solid phase and incubated with increasing concentrations of wt-AGR2 or E60A (monomeric) and Δ45 (dimeric) mutant. Binding was then detected using MAB3.4 and quantified using anti-mouse secondary IgG coupled to peroxidase and chemiluminescence. The data are depicted as binding activity (in relative light units) as a function of increasing AGR2 protein concentration (in triplicate, presented as an average ± standard deviation).

5.2.6 Utilising oligomerisation assays to investigate the use of natural compound libraries to manipulate AGR2 quaternary structure.

Since the development of the ^{2S}MTA and the DSS-crosslinking assay had highlighted its utility for the detection of the oligomeric state of AGR2, it was endeavoured to discover whether a natural or small molecule with bioactivity to modulate the oligomeric stability of AGR2 could be identified. Utilising chemically diverse natural compound libraries (Strathclyde Natural Products Library, Strathclyde Innovations in Drug Research (SIDR), University of Strathclyde) extracted from a broad range of plant families, a high-throughput screen to identify potential regulators of AGR2 was devised. Initially, 70 crude plant lysates from SIDR library 0900.004D3 were prepared into 35 cocktails containing 200 µM of each library condition (final concentration of 400 µM of additive cocktail) and incubated with 500 pM AGR2. These were then subject to the capture by unlabelled MAB3.4 and detection of DyL800-MAB3.4 in the ^{2S}MTA (Figure 5-17A). Raw data was compared to a DMSO-only control and deviation tested for significance using a two-tailed unpaired T-test at 95% confidence interval (a two-tailed test was used since any deviation from the control, either stabilisation of dissociation of the dimer, were of interest). The data suggests that none of the extracted plant lysates could significantly stimulate the formation of AGR2 higher order quaternary structures. Interestingly, however two conditions, containing A4/B4 and C3/D3, presented disruption of the AGR2 oligomer sufficiently to meet the constraints of the statistical analysis.

To identify which of the components of the cocktails were responsible for the bioactive disruption of the AGR2 oligomer, the four possible positive conditions (A4, B4, C3 and D3) were titrated (200, 100 and 50 µM) into the ^{2S}MTA experimental set up. This allowed the identification of the plant extract for further investigation (Figure 5-17B). Compared to DMSO, extracts A4 and C3 did not show any deviation of the oligomer detection from the control; by contrast B4 and D3 exhibited a quaternary structure destabilising activity. Interestingly, extract D3 demonstrated this effect in a dose-dependent manner, where the higher concentrations of compound considerably destabilised the dimer. At the lowest

concentration (50 µM) this disruption was no longer deemed significant. Somewhat confusingly, compound B4 also exhibited a reduction in detection of the oligomer, yet only passed statistical testing at the upper (200 µM) and lower (50 µM) concentrations of the additive, and not at the intermediate (100 µM). What this experiment did confirm was that B4 and D3 were the bioactive components of the oligomer destabilising cocktails identified. For specificity, a further 80 compounds from an additional library (0299.009D3) were tested and compared to the inhibition of the previously identified B4 and D3 conditions of 0900.004D3 library and the inhibitory self-peptide 5 (Figure 5-11) from previous investigations. (Figure 5-17C). Of those tested, none exhibited any stimulation or disruption of the AGR2 oligomeric complexes.

The bioactive compounds which indicated an allosteric modulation of the AGR2 quaternary structure were then compared to the previously discovered modulating self-peptides of AGR2 at 200 µM (Figure 5-18). As expected the stimulatory self-peptide 3 of AGR2 was the only additive which significantly stimulated AGR2 oligomer detection compared to the DMSO control. Disruptive peptide 5, along with the identified natural compound disrupters B4 and D3 exhibited significant reduction in the amount of oligomeric AGR2 detected compared to both the DMSO control and the A4 (non-bioactive) negative control.

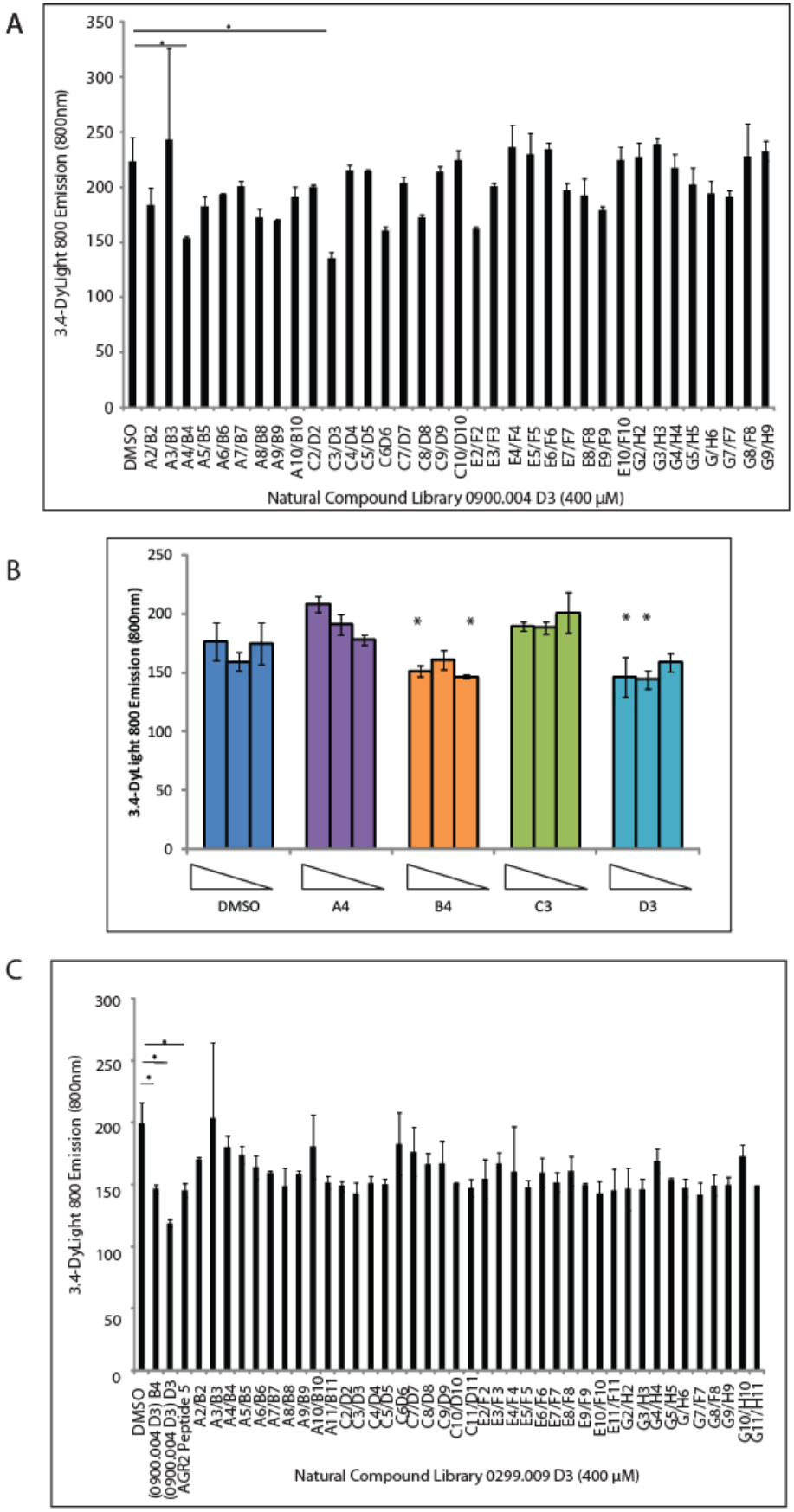

A
3.4-DyLight 800 Emission (800nm)
Natural Compound Library 0900.004 D3 (400 µM)
B
3.4-DyLight 800 Emission (800nm)
DMSO
A4
B4
C3
D3
C
3.4-DyLight 800 Emission (800nm)
Natural Compound Library 0299.009 D3 (400 µM)

Figure 5-17 Application of the two site microtiter assay in the identification bioactive compounds influencing the stability of the AGR2 oligomer using a natural compound library of plant extracts. (A) Testing of 70 compounds of library 0900.004D3 in 400 µM cocktails of doublets of conditions by ^{2S}MTA. Compounds were pre-incubated with recombinant AGR2 for one hour before sandwich capture on MAB3.4 and detection by DyL800-MAB3.4. Significance was tested by a two-tailed unpaired T-test (p-value < 0.05). (B) Individual titrated compounds, A4, B4, C3 and D3, from the positive cocktails (from Figure 5-17A) at 200, 100 and 50 µM tested by ^{2S}MTA. Tested for significance with a one-tailed (less than DMSO control) unpaired T-test at 5%. (C) Testing of a second natural compound library 0299.009D3 compared to the oligomer disruption by AGR2 self-peptide 5 and conditions B4 and D3 from 0900.004D3, significance was tested two-tailed unpaired T-test (p-value < 0.05) as above.

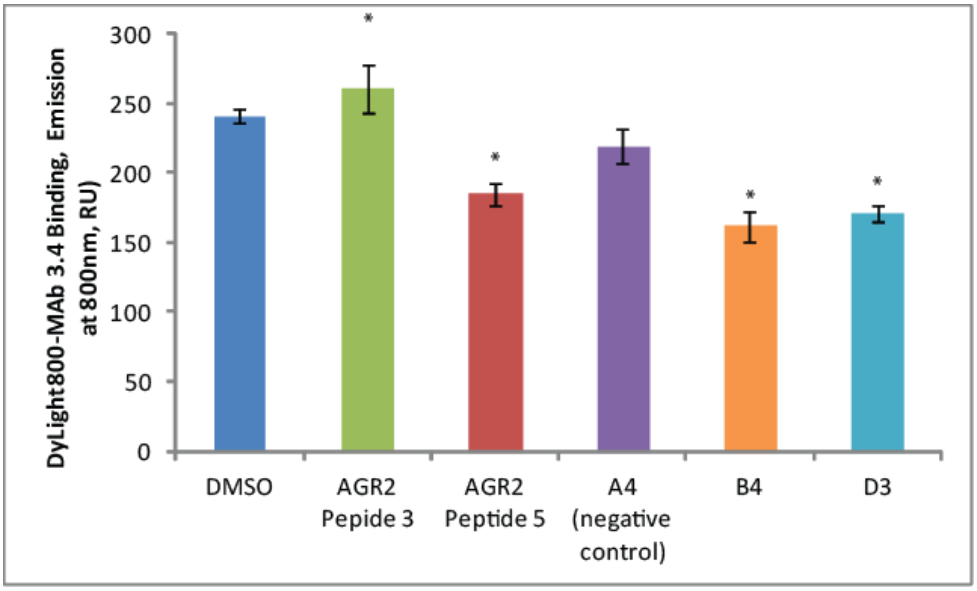


Figure 5-18 Comparison of the level of oligomer allosteric modulation detected by two site microtiter assay of AGR2 by using self-peptide 3 (stimulatory), peptide 5 (disrupting) compared with natural product lysates A4 (negative control), B4 and D3 (both disrupting) from library 0900.004D3. Significance was tested by tested two-tailed unpaired T-test (p-value < 0.05).

As a further validation, the DSS-crosslinking assay was subsequently applied to AGR2 protein (500 pM) incubated with titrated (400, 200 and 100 µM) additions of the positive crude natural compounds (B4 and D3) and a non-bioactive control (A4). Crosslinking at 50 µM was carried out for 30 minutes prior to SDS-PAGE separation and immunoblotting with MAB3.4 (Figure 5-19). The bioactive plant lysates appear to significantly reduce the crosslinking and detection of the dimeric and oligomeric structures of AGR2 compared to the non-interacting control (A4). Compound D3 appears to diminish the monomer to the greatest extent suggesting it exhibits the greatest affinity for the disruption of the AGR2 dimer.

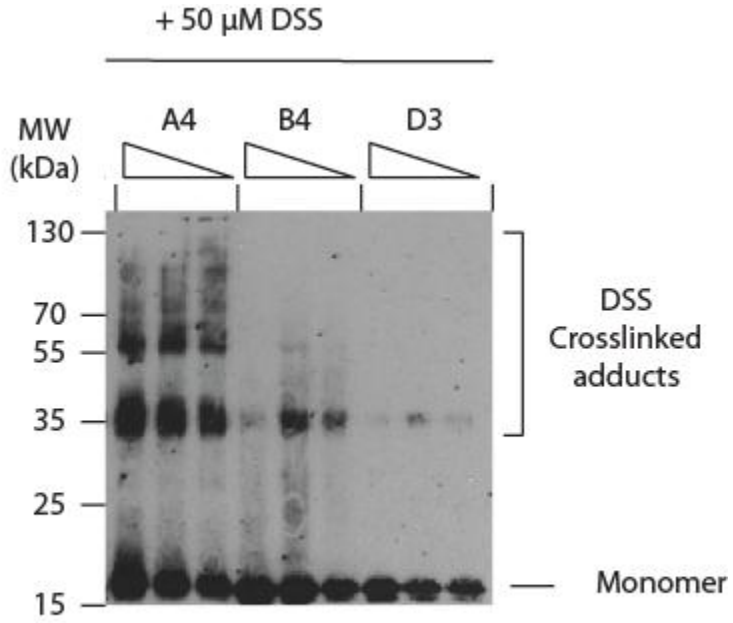


Figure 5-19 DSS induced crosslinking of AGR2 in the presence of natural product compounds determining the effect of bioactive natural products on the oligomeric status of AGR2. AGR2 protein (500 pM) was incubated with 400, 200 and 100 µM natural products A4 (negative control) and B4 and D3 (oligomer disrupters) for one hour prior to 30 minute incubation with 50 µM DSS. Samples where then separated by SDS-PAGE and immunoblotted with MAB3.4 and a peroxidase conjugated rabbit anti-mouse secondary antibody and chemiluminescence.

Since library compound D3 was determined as the most bioactive oligomer/dimer disrupter of AGR2, this crude lysate was then fractionated by high performance liquid chromatography into 4 fractions containing active molecules. These fractions (7, 8, 9 and 10) were again subject to the two oligomerisation assays, ^{2S}MTA and DSS-crosslinking assay, to identify which fractions contained the bioactive molecule for disruption of AGR2 quaternary structure. The ^{2S}MTA (Figure 5-20A) indicated that fractions 7 and 10 exhibited the greatest level of dimer disruption compared to DMSO. Fractions 7 and 10 of D3, at the greatest concentration (400 µM), also demonstrated oligomer disruption greater than the positive control crude lysate D3. Fractions 8 and 9 did not appear to have any effect on the oligomer detection, with a nature similar to the DMSO control. For further investigation of fractions 7 and 10, they were subject to the DSS-crosslinking assay (Figure 5-20B). In this assay, the disruption of the oligomer by fraction 7 was not apparent, and the level of dimerisation was comparable to the DMSO control and greater than the fraction 8 non-bioactive control. Fraction 10, however, demonstrated a dose-dependent disruption of the AGR2 dimer, by both assays and thus is the prime candidate for further work and structural identification.

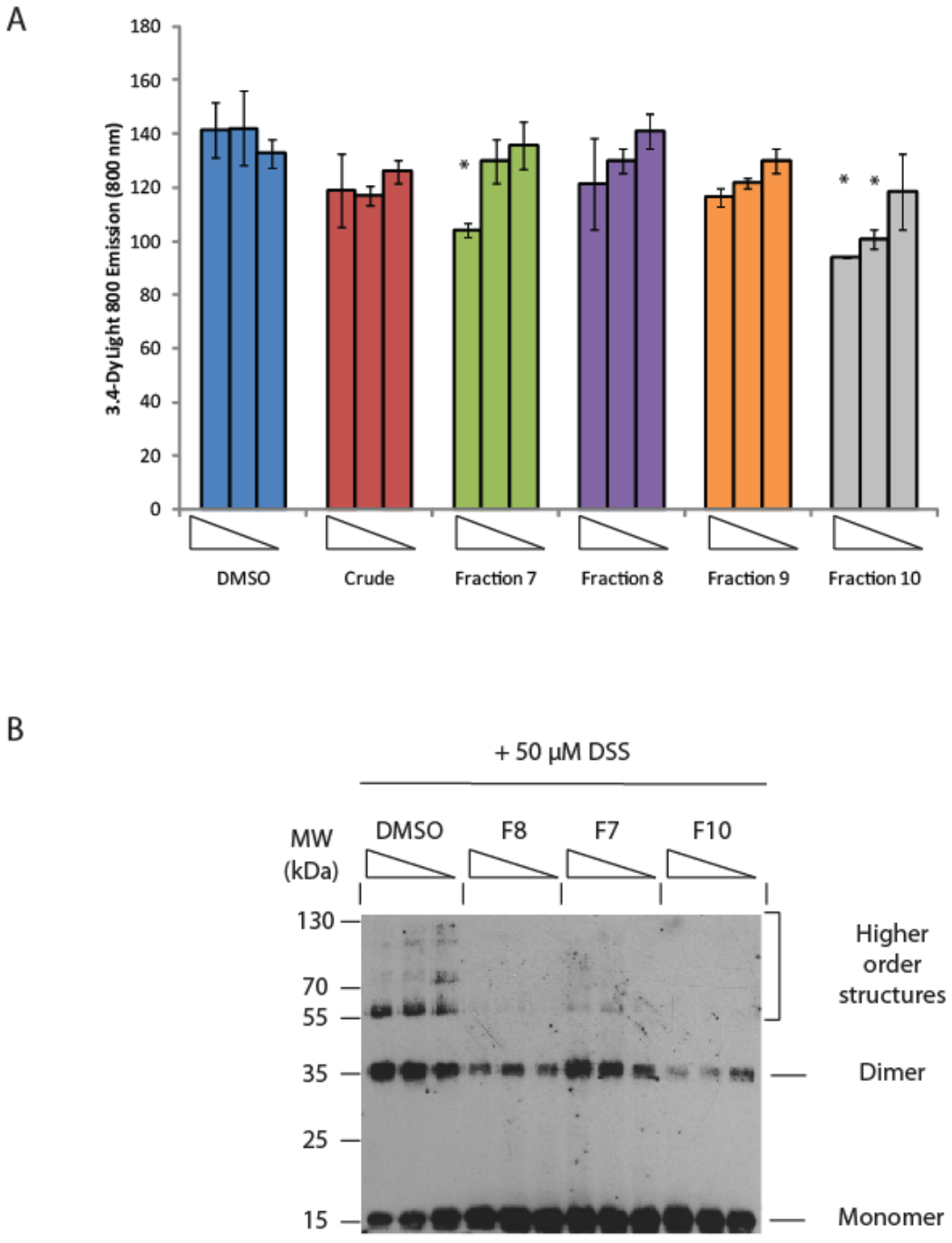


Figure 5-20 Identification of the bioactive fraction of SIDR compound D3 in the disruption of AGR2 oligomeric complexes. Testing of high performance liquid chromatography separated fractions of compound D3 on the oligomerisation of AGR2 by (A) two site microtiter assay at 50, 25 and 12.5 µg of fractions with 100 ng of AGR2 and detection by DyL800-MAB3.4. and (B) DSS-crosslinking assay with 50 µM DSS incubation for 30 minutes. Detection was by imuunoblot by MAB3.4 and peroxidase conjugated secondary antibody and chemiluminescence. Significance was tested by tested one-tailed unpaired T-test (p-value < 0.05).

5.3 Discussion

Understanding the properties of AGR2 at the molecular level has become a key aim of research to understand the bioactivity of this novel oncoprotein. Until recently, the majority of structural knowledge of AGR2 was through interactomic studies which had identified, using yeast-2-hybrid (169;199) and co-immunoprecipitation (172) studies, a number of interaction partners of AGR2. Novel research methods however identified a homo-dimeric structure of AGR2, putatively through two complementary interfaces, the 60-EALYK-64 interface (201), or an intermolecular disulphide interaction between cysteine-81 (202;203). Our initial biochemical characterisation, prior to the publication of the above reports, by size exclusion chromatography indicated that recombinant $AGR2_{21\text{-}175}$ protein exhibited an apparent mass greater than that expected of the monomeric subunit on molecular sizing columns, and was more consistent with a homodimeric quaternary structure of the protein. In contrast to the publication by Patel *et al*. (201), initially we resisted the truncation of the protein to remove the unstructured N-terminal 45 amino acid residues of AGR2. *In vivo,* following the cleavage of the signal peptide recognition sequence (amino acid residues 1-20), 25 remaining residues of this disordered motif and may contribute to the physiological instability induced by protein disorder, and this is essential for thorough biochemical characterisation, particularly with respect to the establishment of monomer-dimer equilibrium. Subsequently, the purification of the homodimeric protein retaining the affinity for TxIYY binding (187;275), it was hypothesised that the quaternary structure may influence, or be essential for the biochemical functions of AGR2, as is seen in the recent publication regarding BiP/GRP78 interaction (202). Therefore, the objective was to create an assay which could be used to determine the oligomeric state of the protein of interest, and utilise this assay in the creation of mutants with altered oligomeric equilibrium, and to drive the application of natural compound libraries to uncover small molecules which could modulate the balance of monomer-dimer.

Monoclonal antibodies derived against the open reading frame of AGR3 protein, were tested for cross reactivity with AGR2 (Figure 5-5 and 5-6) and MAB3.4 was found to exhibit a strong affinity for AGR2 (248). The epitope for this antibody could be mapped to a linear site in AGR2 protein (PLMxI, Figure 5-6A and 5-6D), unlike the commercially available AGR2 monoclonal antibody (Figure 5-6B), thus the antibody could be used for specific detection of this epitope in biochemical assays. Thorough characterisation of the antibody was necessary to ensure selectivity of the defined epitope for downstream studies.

The assay devised measures oligomer stability via a two-site microtiter assay, or modified sandwich assay, in which an oligomeric protein can be captured in the solid phase and detected in small volume liquid phase using a well characterised monoclonal antibody that binds to the same linear epitope. Therefore, this assay cannot detect monomeric species, but rather detect the level of protein oligomerisation (Figure 5-7). Assays targeted at deciphering oligomer formation using a similar principle have previously been reported in the Alzheimer's research field studying the formation of amyloid plaques (395;396). Prior to applications of the assay in the analysis of bioactive molecules, the assay was validated to ensure that the DyL800 labelling strategy did not detrimentally affect epitope recognition, that following conjugation the DyL800 fluorophore was functional (with 750 nm excitation and 800 nm emission) and was detectable using the 800 nm channel of the Licor Odyssey, and that the capture-detection hypothesis using the same monoclonal antibody allowed capture and detection of the oligomeric protein. Therefore, a range of titrations of the detection antibody were tested in preliminary studies to ensure that the optimal antibody concentration was being used; enough DyL800-labelled antibody to allow detection but not excessive amounts which would compete with the unlabelled-MAB3.4 capture on the substrate. This also allowed the appreciation of association and dissociation constants (on- and off-rates) of the affinity of the antibody, which could be affected by incubation time, concentration of oligomeric protein and concentration of detection antibody. A comparable assay to this involves the use of Surface Enhanced Raman Spectroscopy (SERS) in an aggregation based assay, which measures the oligomerisation of the ubiquitin ligase MDM2 through the

congregation of Raman active particles conjugated to MDM2 to provide a read-out of oligomerisation (397). Similarly, this Raman based assay was used in competition to identify self-peptides from MDM2 which interfere with oligomerisation and attenuate the SERS signal. Both SERS and ^{2S}MTA described here have advantages over each other. The main advantage of SERS is that biochemical effects can be measured in real-time, unlike ^{2S}MTA where incubation to allow the interaction of antibody and epitope is required. However, ^{2S}MTA can detect oligomerisation in the absence of any ligand thus this assay measures the quaternary structure more directly. ^{2S}MTA was able to detect attenuation of AGR2 oligomerisation using self-peptides from the published dimer interface from amino acids 60-EALYK-64 (201). Amino acids outwith this motif also attenuated oligomer stability (e.g. peptide 4, Figure 5-11), suggesting that the dimer motif requires the stability of adjacent peptide chains to allow the formation of the homodimer. Surprisingly, peptides were also identified that could function *in trans* to stimulate the oligomerisation of AGR2 in the ^{2S}MTA. These peptides are derived from the intrinsically disordered region of the N-terminal domain of AGR2 (Figure 1-5) and have the ability to act as positive effectors of oligomerisation suggesting they can disrupt the function of the natural N-terminal sequence *in cis*. There is a growing field in the understanding and use of and function of unstructured peptide motifs in driving the weak, regulatory, but highly specific protein-protein interactions so necessary and important in cell signalling (376). Interesting comparisons can also be made with peptides and mimetics which can allosterically stimulate protein-protein interactions (398), and whether AGR2 dimerisation/oligomerisation stimulatory peptides have any influence on the manipulation of AGR2 biochemical function remains to be elucidated.

Using the ^{2S}MTA as a rapid screen for oligomerisation capacity in whole proteins, coupled to the data from self-peptide disruption of the oligomerisation equilibrium, a site directed mutation strategy targeting the bulky, hydrophobic or charged residues of peptide 5, thought to be essential in the formation of the dimer interface (Y58, E59 and E60), were subject to alanine mutation. This confirmed that the mutation of glutamic acid residue 60 to alanine resulted in a protein species with a quaternary structure equilibrium skewed toward the monomeric form (Figure 5-14). This feeds

into the concurrent publication that glutamate-60 forms an intermolecular salt bridge interaction with lysine-64 of the adjacent interacting partner (201). This could be confirmed using the DSS crosslinking assay (Figure 5-15), which allows covalent binding of lysine residues within an 11.4Å proximity. The AGR2-E60A mutant was less readily available than the wild-type protein for DSS-induced chemical crosslinking, suggesting that this mutant exhibited a spatial distribution of protein molecules distinct to that of wt-AGR2. As such, the key crosslinked site between lysine-95 residues, previously mapped by mass spectrometry (203), were not within the necessary distance of the crosslinking spacer, thus AGR2-E60A exhibited an attenuated dimer conformation. By contrast, the truncation of the N-terminal disordered region of the mature AGR2 protein, residues 21-45, as suggested by the self-peptide 2 and 3 stabilisation of recombinant AGR2 protein (Figure 5-11 and 5-14), resulted in detection of increased oligomerisation by ^{2S}MTA. This conclusion was supported by the supplementary DSS crosslinking assay (Figure 5-15C), and DMSO stabilisation of AGR2-Δ45 (Figure 5-15D), which indicated that AGR2-Δ45 exhibits a dimer with significantly increased stability. The role of intrinsically disordered regions of proteins involved in protein-protein signalling has recently come to light, and a significant level of research effort is driving the understanding of these motifs of complex function (399-402). Over the past decade, this has driven somewhat of a paradigm shift away from proteins adopting highly ordered structures and interacting through a 'lock and key' mechanism (403), where the interacting proteins must exhibit the correct conformation before the formation of the interaction interface. Instead, due to the observation of a significant number of proteins exhibiting numerous conformations, and are likely to have some fluidity in secondary, tertiary and quaternary structure to allow a mechanism of interaction termed 'induced fit' (404;405), where the structure of interacting motifs are malleable thus can form the most stable interaction of lowest energy state. There is a growing body of evidence of intrinsic disorder found in cell signalling proteins and transcription factors, suggesting an important role in their regulatory capacity (406). Indeed from structural studies of AGR2 it has been uncovered that deletion of amino acids 21-40 considerably increases the stability of the AGR2 dimer by as much as 3 orders of magnitude (201). Indeed, when comparing the greatest concentrations of

DSS-crosslinking (Figure 5-15A, 5-15B and 5-15C, far right lanes), both wt-AGR2 and AGR2-E60A exhibit higher molecular mass adducts which are not prevalent in the crosslinking of the AGR2-Δ45 mutant. This may be due to the enhanced instability of the former two proteins due to the presence of the disordered N-terminal domain that might affect the protein structure, whereas AGR2-Δ45 may be more rigid and resist such non-specific higher molecular mass species. Taken collectively, these data suggest that the disordered N-terminus of AGR2 may form a regulatory *cis*-acting element allowing modulation of the dimer status.

Following the cloning of AGR2 constructs with stabilised and depleted dimerisation, it was important to determine whether the contribution of the quaternary structure is important for known bioactive properties of AGR2. In small volume, microtiter assays, the mutants were tested to identify whether dimerisation was required for the interaction with the TxIYY peptide motif (187;275) and the validated interaction with Reptin (199) (Figure 5-16). The TxIYY peptide was identified in a peptide-phage display interaction screen as a highly specific peptide ligand that can interact with AGR2 *in vitro* (275) and *in vivo* (187) and presumably reflects an intrinsic peptide docking motif on its client proteins. Interestingly, the AGR2 monomeric mutant, E60A, bound as stably as wild-type AGR2 or the enhanced dimer mutant, AGR2-Δ45. This data indicates that if the TxIYY docking site represents a cellular binding interface for AGR2 client proteins, a study which is still underway (unpublished data), then the monomeric form of AGR2 has the potential to be equally active in this assay. Contrastingly, the binding of AGR2 to Reptin was stabilised by the creation of the enhanced dimer protein (AGR2-Δ45) and attenuated by the monomeric E60A mutant. Thus the determinant in AGR2 that interacts with Reptin (primarily within amino acids 104-111 (199)) exploits the dimeric subunit structure of AGR2 protein for Reptin interactions. Solving the interface of AGR2-Reptin may facilitate the development of biologics that disrupt this protein interaction to determine how it might contribute to cancer cell growth. With reflection of a recent publication on the functional role of dimeric AGR2 in the attenuation of endoplasmic reticulum induced cell death (202), the interaction of dimeric AGR2 with Reptin mirrors that of the AGR2-BiP/GRP78 interaction. Both

require AGR2 to be in a dimeric structure for the interaction to take place. The study presented herein and this publication (202), use different mutant constructs to reach the conclusions of the functional role of the dimer, however both provide evidence of the disruption of the dimer, suggesting that both the cysteine-81 residue (202;203) and the 60-EALYK-64 motif (201) act in a complementary fashion to drive the interaction of the AGR2 subunits, and are not mutually exclusive.

Since we now know that the dimerisation of AGR2 is necessary for at least some of the biochemical functions of the proteins, with particular reference to the oncogenic properties of AGR2, it was endeavoured to undertake a methodology to provide a natural compound lead for potential drug development studies. Natural products have been described as the single most productive source of leads for the development of drugs (407) and in particular cancer drug discovery (408). Thought, at least in part, to be as a result of the chemical properties of small molecule natural products found that the majority of successful compounds developed into drugs are closely complaint with Lipinski's Rule of Five and thus more pharmacologically suitable for therapeutics than synthetic drugs (409). Novel approaches for natural product screens begin with crude extracts of plant or microbial lysates, resulting in a complex mixture of perhaps several hundred different compounds (410) which can then be applied as a crude cocktail in biochemical assays, removing the need for expensive, intensive synthesis to purify a library of pure small molecules for screening, instead the natural product extracts can be used early in the screening process (Figure 5-17, 5-18 and 5-19). Once bioactive lysates have been identified, and validated by a secondary assay (and compared to a positive control, e.g. the self-peptides of AGR2 which modulate dimerisation/oligomerisation), these crude lysates can be fractionated into simplified reagents for further screening (Figure 5-20). Presented here is that fraction 10, from compound D3 of 0900.004D3 from the academically available SIDR library (University of Strathclyde), contains an active molecule which can dissociate the AGR2 dimer to a statistically significant effect. Further studies are required to identify the bioactive small molecule within this fraction, by techniques such as mass spectrometry. Subsequently, this compound could be purified or chemically synthesised for bioactivity in further assays, such as

thermal denaturation assays (TDA) to determine whether the molecule drives a shift in the melting temperature, and thus structure of the AGR2 protein. Subsequent NMR or crystallography studies would allow structural mapping of the interaction, and downstream optimisation of the chemical structure of the compound for increased efficacy. Finally, biomolecular *in vitro* and *in vivo* assays should be exploited to determine whether the bioactive compound can influence the interaction of AGR2 with Reptin or BiP/GRP78. This would provide information on the effectiveness of AGR2 oligomer targeting in a drug development scenario, for example in the reactivation of p53.

In conclusion, presented here is the development of a microtiter based assay that can be used to measure the oligomerisation state of AGR2 in small volumes, utilised in the study of motifs necessary for the stabilisation and destabilisation of the AGR2 oligomer. Such an assay was used to identify ligands, like self-peptides, that can allosterically regulate positively multimer stability thus forming a proof-of-concept assay for screening for cellular proteins or ligands that might be natural regulators of AGR2 dimer stability. Using this knowledge we identified a biochemical function of the dimeric protein in the interaction with Reptin, and subsequently begun a drug discovery program to utilise natural compound libraries in the disruption of the AGR2 quaternary structure.

Chapter 6: Final Conclusions and Future Perspectives

Aberrations in the function of p53 is one of the key outlying features of many manifestations of human cancer. p53 is a central hub protein, collaborating with numerous interaction partners, to protect the cell from propagating erroneous genes and suppressing tumour formation by controlling cell cycle and proliferation. The majority of these dysfunctional phenotypes are as a result of the expression of a mutant form of the p53 hub protein, yet a significant number of cancer types exhibit the wild-type p53 gene product, resulting in p53-dependent activity being modulated at the post-transcription level. Thus, the reactivation of wild-type p53 has become a key strategy targeted by anti-cancer therapeutic programs. The discovery that the overexpression of AGR2 is apparent in a significant number of cancers, its expression predicts poor prognosis, is linked to resistance of a number of current chemotherapeutic compounds, and has a suppressive effect on the p53 response to DNA damage has driven research investment in this field as a potential drug target for the reactivation of p53. Subsequently, AGR2 has been found to stimulate cancer cell proliferation, invasion and survival, influence chemotherapeutic resistance in *in vivo* models and highlighted as a pro-metastatic factor in the progression of disease. Yet, prior to this study, little was known about the molecular arrangement, regulation, subcellular function or signalling pathway components of the protein.

The data presented herein proposes a reliable *in vitro* model of AGR2 expression which exhibits several of the published functions of AGR2 in cancer cells including the enhanced the unfolded protein response to endoplasmic reticulum stress, appropriate subcellular distribution as a result of the non-canonical ER retention motif and significant proliferative advantage over an isogenic cell line lacking the subtle constitutive expression of low levels of AGR2. Further, a panel of cell lines were developed for the interrogation of the function of AGR2 with modified subcellular distribution to unravel the hypothesis of AGR2 as an autocrine/paracrine signalling factor, and as an ER-localised protein disulphide isomerase molecular chaperone, together with an isogenic cell line expressing a closely related gene of interest, (AGR3). These models demonstrated that wild-type AGR2 overexpression

did not have a significant consequence on the transcriptional landscape of the cell, suggesting that any pro-oncogenic activity may manifest at an alternative stage, such as at the protein level. It may be pre-emptive to suggest that AGR2 does not have any effect on the transcriptional machinery of the cell, as these cells were engineered to express AGR2 protein below the endogenous levels seen in AGR2-positive cancer cell lines, but it may highlight that the dominant effect of AGR2 overexpression is not expected at the mRNA stage. To investigate the transcriptional effect of AGR2 further, for example to identify misregulated p53-dependent gene targets, models expressing greater levels of AGR2 are required. This cell panel however does provide an excellent representation for future studies requiring an *in vitro* model of AGR2 expression since the mechanism of gene recombination and expression levels of gene of interest mean that this is a more reliable and reproducible methodology than the more frequently used transient transfection.

In addition to the development and validation of the experimental cell line panel, a proteomics approach was utilised in the application of mass-spectrometry based quantitative protein expression methods to assess the effect of low-level, stable expression of AGR2 at the protein level. This study aimed to define a mechanism for how the emerging oncoprotein, AGR2, can remodel the cancer landscape. Using blind, unbiased bioinformatical data analysis these data confirmed that expression of wild-type AGR2 protein in this cell line had a suppressive effect on the transcriptional activity of p53, and this could be validated using immunoblot and gene silencing methods in additional endogenous AGR2 expressing cell lines. This suppression of p53-dependent activity was further supported by the upregulation of a key protein marker, used in immunohistochemistry of cell proliferation, Ki-67. Further, data-driven studies identified the expression of the previously published p53 inhibitor, TSG101, as the outlying most misregulated protein expression level following AGR2 introduction. Subsequent manipulation of TSG101 levels in a range of cell lines indeed confirmed the negative correlation between TSG101 protein expression with p53 levels and p53-dependent activity. Therefore, a model is proposed where AGR2 presence drives the expression of TSG101 to negatively regulate the expression and activity of p53. Further immunohistochemical studies

are required to assess the co-expression of AGR2 and TSG101 in human tumour biopsies, allowing analysis of whether TSG101-mediated suppression of p53 is an observable trait. The next steps of this study are to define the mechanism whereby AGR2, predominantly an ER-resident molecular chaperone, promotes the expression of TSG101. It is hypothesised that this may be through the appropriate maturation of nascent TSG101 gene products, however based on the non-exclusive ER localisation of AGR2 it is possible that TSG101 is stabilised in a non-ER related manner. Further, it is essential to understand how TSG101, an endosomal trafficking associated protein influences the activity and expression of the nuclear transcription factor, p53. Previous studies in the AGR2 field have proposed that transient AGR2 expression results in the nuclear exclusion of p53 in the cytoplasm, and as a result attenuated p53-dependent transcriptional activity. TSG101 may perform some role in the re-localisation or compartmentalisation of p53, or alternatively other studies suggest that TSG101 participates with MDM2 in an autoregulatory loop to modulate cellular levels of both proteins, and of p53, by affecting protein decay. On that basis, there are several hypotheses to challenge to uncover the mechanism of AGR2-TSG101-p53 signalling. The elucidation of a mechanism of this signalling axis is the key point in defining whether this pathway provides a possible target for therapeutic intervention allowing the reactivation of wild type p53 activity.

Along with the pro-oncogenic signalling of AGR2, a molecular interrogation screen was devised, engineered and optimised to allow analysis of the oligomeric status of AGR2 protein. This oligomerisation assay allowed the examination of AGR2 as a dimeric unit and the interface forming cofactors, the requirement of the dimeric structure for Reptin interaction and the first study of potential AGR2 dimer stabilising or disrupting bioactive compounds. The dimeric arrangement was concurrently uncovered by several groups during the scope of this study, each with a variation on the proposal of the amino acid determinants required for dimer formation. This study, using an overlapping self-peptide library, concurred that dimeric AGR2 required the glutamate residue 60 to form an intermolecular salt bridge interaction to maintain the structure, and in addition brought adjacent lysine-95 molecules in close enough proximity to be crosslinked by DSS. The study also

independently suggests that the unstructured N-terminal domain, cleaved in previous studies to aid structural determination, plays a destabilising role in the formation of the quaternary structure. Utilising a natural product library screen, presented herein is the first suggestion of a bioactive compound which can disrupt the monomer-dimer equilibrium of AGR2. The essential progressions of these results are to analyse the significance of the mutated proteins with modulated quaternary equilibrium, and the bioactive compounds, on the activity of AGR2 *in vitro* and *in vivo*. To this end, in the subsequent chapter 7-1, preliminary data has been collected regarding the development of a crystal structure of native AGR2. Further optimisation of crystallisation conditions, and resultant high resolution structure of AGR2, will allow the soaking in of potential bioactive compounds followed by characterisation of the resultant interacting complex. Subsequently, the global structural changes required to promote a shift in the equilibrium of the monomer-dimer balance can then be defined. These follow up studies will begin to provide structural assessment of the AGR2 dimer/monomer, and how destabilisation will affect characteristics such as interaction partner binding. What remains unclear is what role the monomer-dimer equilibrium plays in the normal and pro-oncogenic properties of AGR2 (Figure 6-1). Utilising the Flp-In methodology of low level stable gene expression, an experimental cell line expressing AGR2-E60A and AGR2-Δ45 mutants should be constructed for comparison with the normal function (CHOP response to ER stress) and pro-oncogenic properties of the wild-type gene in assays presented in this study, such as cell growth advantage over AGR2-null cells and p53-inhibition through response to cisplatin damage. As a result, the biomolecular function of the monomer-dimer equilibrium could begin to be understood, and investigated whether manipulating this balance influences the oncogenic properties of AGR2 and as a result could be targeted for novel anti-cancer therapeutics.

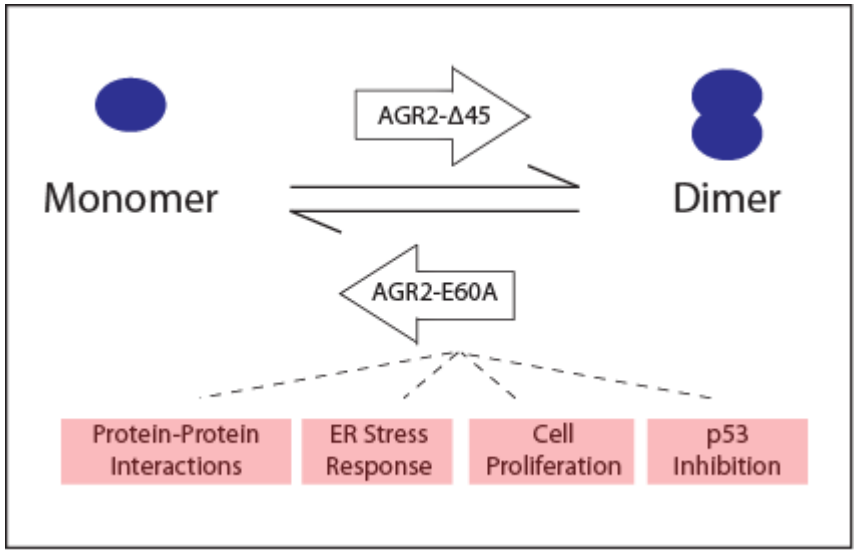


Figure 6-1 Schematic of AGR2 monomer-dimer equilibrium and tools to investigate the role of equilibrium shift on the *in vivo* functions of AGR2.

Chapter 7: Preliminary Data

7.1 Developing optimal conditions for AGR2 protein crystallisation

The structural characterisation of the wild-type AGR2 protein is essential for further understanding of function, and further for the development of novel, high affinity interacting compounds which may allow attenuation of activity in human diseases. X-ray diffraction studies are useful in revealing information about the dynamics and heterogeneity of protein structures, through the highly accurate determination of well-refined, high resolution crystal structures (411). More than 85% of the macromolecular structures currently in the Protein Data Bank (PDB) have been determined by X-ray diffraction, making it the most successful method for determining structures of large molecules (412). The vast majority of protein structures determined to date curated in the PDB are at medium to high resolution (2.8 to 1.5 Å) (411). This resolution is highly important since it provides greater accuracy sub-2Å (413), allowing a reduction in errors of atomic positions and reduces misinterpretation of structure. Accurate atomic positions allow optimal evaluation of hydrogen bonding interactions and subsequent derivation of essential stabilising interactions.

In order to begin to develop a screen for crystallisation conditions for AGR2, several conditions were required to be met: (i) Protein sample must be pure and free from contaminants; (ii) active in a functional assay; and (iii) sample must be homogeneous, chemically (free from proteolytic fragments) and conformationally (oligomerisation, aggregation, denaturation). From Chapter 5, we can confirm that the sample is indeed pure and can be synthesised and purified in large quantities (Figure 5-2). The protein also demonstrates a functional activity in binding to the previously published AGR2 interacting aptamer (Figure 5-3) (187;275). However, the sample is not fully homogeneous, with populations exhibiting different

oligomerisation statuses evident by size exclusion chromatography (Figure 5-4). At high concentrations (>2.5 mg/mL), AGR2 exists predominantly as a dimeric species as calculated from the size exclusion chromatography (Figure 5-4B and 5-4C). While not being most favourable, this could be overcome through optimisation of crystallisation and buffer conditions to assist in homogenous crystal packing.

The method used for crystallisation was the most commonly used method termed vapour diffusion. The protein solution (5-20 mg/mL) is prepared in a hanging or sitting drop that equilibrates against a reservoir containing crystallising agents such as precipitants (polyethylene glycol, PEG) or additives, at conditions greater or less the concentration in the droplet, in a system closed from the external environment. Initially, the droplet of protein solution contains an insufficient concentration of precipitant for crystallisation, over time water vaporises from the drop and transfers to the reservoir (mother liquor) and the precipitant concentration increases to a level optimal for crystallisation. The protein solution can exist in 4 different conditions: (i) precipitated, where the protein aggregates into a non-uniform structure, (ii) nucleation zone, where the protein begins to form needles, micro-crystals or nucleates but most of the protein is still in solution, (iii) the metastable zone of crystals, the goal condition or (iv) under-saturation, where the protein remains in solution (414). The aim of vapour diffusion is to prepare a solution where the solution is initially under-saturated, that the protein is in solution, and as the water diffuses out of the droplet, the concentration of precipitants increase gradually, and the protein solution moves to the metastable zone, resulting in the formation of compact crystals. The inclusion of additive molecules (such as ammonium sulphate) provides minute variations in the physiological conditions, whether these vary the intermolecular contacts between macromolecules or eliminate intermolecular interactions, to drive crystal nucleation allowing stabilisation and promotion of crystal lattice formation. Vapour diffusion format lends well to multi-well plate arrangement where conditions such as precipitant concentration, additive concentration and pH of the buffer solution can be manipulated in a matrix such that the conditions can be simply monitored (415).

Early results of this study exhibited small crystals, larger than microcrystals, (Figure 7-1) consisting of small nucleated protein products at conditions of 100 mM Tris pH 7.5, 25% PEG4000 supplemented with 5% ammonium sulphate. Following isolation, freezing in liquid nitrogen and X-ray diffraction, this crystal diffracted to 15 Å, suggesting that molecules are not closely enough packed into a crystal lattice such that X-ray diffraction provides information at significant resolution. Further precise modification and analysis of these conditions did not exhibit significant improvement on the resolution of the crystals thus further optimisation strategies were sought.

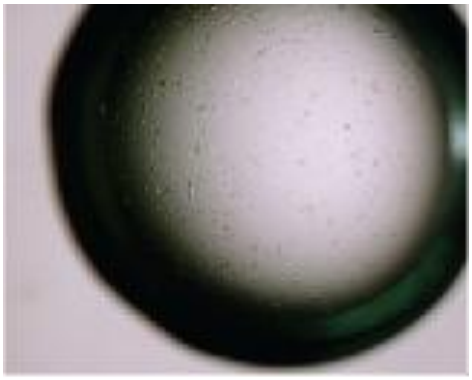


Figure 7-1 Microcrystal formation of AGR2. Mother liquor was comprised of 100 mM Tris pH 7.5, 25% PEG4000 and 5% ammonium sulphate. Protein was concentrated to 10 mg/mL and 1.5 μL was mixed with 1.5 μL of well solution in a hanging drop for 5 days. Crystal was isolated, frozen in liquid nitrogen and diffracted using synchrotron X-ray source (Diamond) and diffracted to 15 Å.

Sparse matrix screens were prepared using commercially available kits which incorporate empirical conditions varying a broad range of salts, polymers of different molecular weights, organic solvents and varying pH. Kits used were Morpheus, JCSG+, Clear Strategy Screen 1, Clear Strategy Screen 2 and PGA (all from Molecular Dimensions). Each of these contains 96 pre-formulated conditions in matrices of protein crystallisation conditions which were suitable for sitting drop analysis in a 96 well microplate arrangement. Coupled to the Oryx 8 robot (Douglas Instruments), CrysCam X-Y stage and CrysScore software (Art Robbins Instruments), these could be prepared in a high throughput arrangement and incubated for 3 (Figure 7-2) and 7 days (Figure 7-3), followed by rapid imaging of the 96-well crystallisation plates. The CrysScore software scores the droplet from (low) red to blue (high) based on phase separation, micro-crystal or small crystal formation. Several conditions highlighted in Figure 7-3 were isolated and made into seeding stocks which could be reseeded with protein into the same condition to provide a nucleation point for new crystals, however this proved unsuccessful.

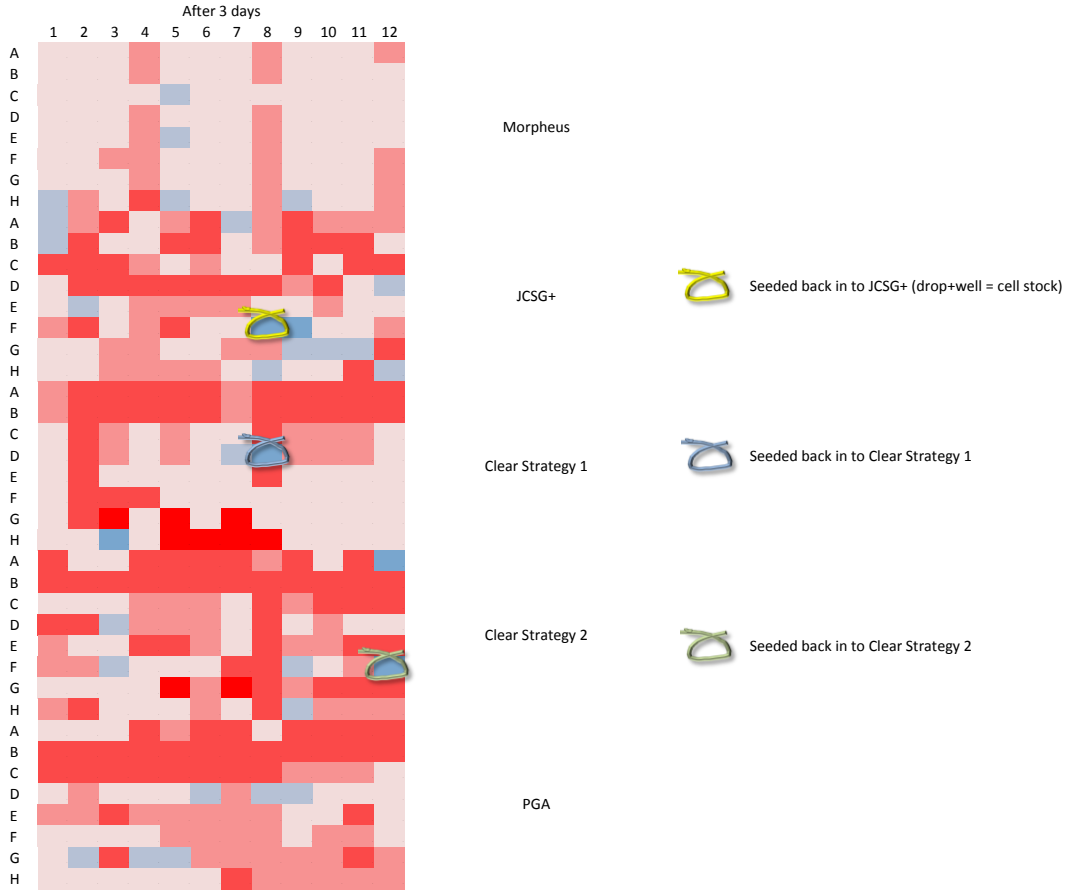


Figure 7-2 Sparse matrix screens using Morpheus, JCSG+, Clear Strategy 1, Clear Strategy 2 and PGA with 48 hour incubation. 100 µL of each condition (Molecular Dynamics). was transferred to the corresponding well in a 96-well microplate. 0.5 µL of 10 mg/mL protein was mixed with 0.5 µL of mother liquor using the Oryx8 robot (Douglas Instruments) and incubated for 3 days. Plates were then screened with CrysCam X-Y stage, and scored using CrysScore software (Art Robbins Instruments). CrysScore ranks conditions based on the phase separation, micro-crystal or small crystal formation of the protein solution. Blue indicates a better droplet for crystal production and red colouration demonstrates conditions less conducive to crystal formation. Conditions circled were removed and provided seeding stocks for downstream analysis.

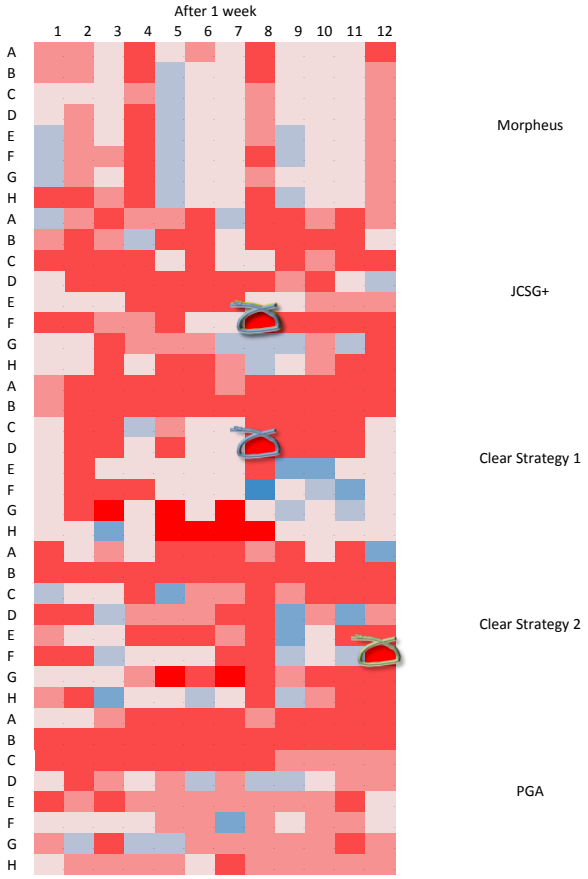


Figure 7-3 Sparse matrix screens using Morpheus, JCSG+, Clear Strategy 1, Clear Strategy 2 and PGA with with 7 days incubation. A heat map was produced from CrysScore analysis packacge. CrysScore ranks conditions based on the phase separation, micro-crystal or small crystal formation of the protein solution. Blue indicates a better droplet for crystal production and red colouration demonstrates conditions less conducive to crystal formation.Circled are the conditions which were previously removed for seeding stocks.

Subsequently, several crystals from these screens were stored in liquid nitrogen prior to X-ray diffraction data collection. Data analysis on the CrysScore derived highest scoring (most blue in Figure 7-3) Clear Strategy 1 condition F8 (Figure 7-4A) disappointingly resulted in limited diffraction and suggested a level of precipitation. However, a poorer scoring crystal from Morpheus condition C3 demonstrated data of better resolution at 5Å (Figure 7-4B). Whilst being a significant improvement on initial studies, this was not sufficient to resolve amino acid side chain arrangements and hydrogen bond interactions (bearing in mind that maximum hydrogen bonds are 3.3Å in length). Literature review (416) identified the Morpheus C3 condition to contain 10% PEG4000, 20% glycerol, 0.03 M of NPS additive (containing $NaNO_3$, Na_2HPO_4 and $(NH_4)_2SO_4$) and 0.1 M MES/Imidazole pH 6.5. Ongoing studies are currently optimising these conditions within a narrow matrix, including seeding techniques, in order to produce crystals which diffract to less than 3Å.

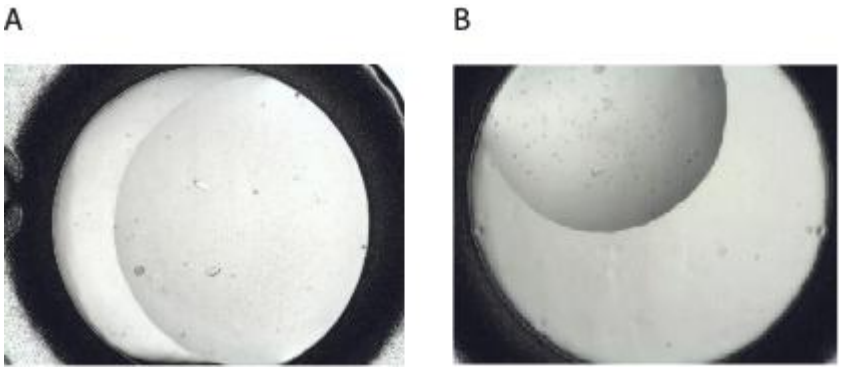


Figure 7-4 Images of crystals derived from the sparse matrix screens . (A) Image of the highest CrysScore scoring condition of Crystal Strategy 1 condition F8, which yielded poor X-ray diffraction data. (B) Image of the Morpheus C3 condition, which while not exhibiting a high score by CrysScore provided diffraction data resolving to 5Å.

During the progress of this study, the NMR solution structure of AGR_{41-175} (2LNS (201)) and the crystal structure of the highly similar AGR3 protein (3PH9, (417)) were published. These are useful in the visualisation of structural conformation of the AGR2 protein; however, the field still lacks the complete structure of the mature $AGR2_{21-175}$, functional unit for analysis of compounds which can be visualised to manipulate the oligomeric structure of AGR2. This preliminary data provides the first steps of optimising the conditions for the reproducible development of AGR2 crystals which allow high resolution X-ray diffraction which can subsequently be used for the development of a highly accurate crystal structure. An accurate crystal structure could then be subject to *in silico* screening for novel interacting compounds (418), hydrogen-deuterium exchange mass spectrometry (419) to analyse the binding interactions with Reptin and other yeast two hybrid protein hits, as well as identifying the interface with the natural compound present in fraction 10 of the D3 condition of the SIBR library (Chapter 5).

7.2 Analysing the effect of AGR2 expression on the profile of secreted proteins

There is a substantial amount of evidence to suggest that AGR2 functions as a pro-metastatic factor in several cancers including breast cancer (420). Transfection of AGR2 cDNA into a benign rat mammary cell line, Rama 37, induced a metastatic phenotype *in vivo*, and these cells injected into the mammary fat pads of syngenic rats had the propensity to develop lung metastasis (219). Over-expression of AGR2 was also identified in metastatic hepatocellular carcinoma (HCC), and an increase in AGR2 expression increased invasion of HCC cell *in vitro* and *in vivo* (208). While in prostate cancer, over expression is reported in metastatic tumours (246), such that AGR2 is being explored as a potential urine- (245) and blood- (259) based diagnostic marker of metastatic prostate cancer. What remains unclear is how AGR2 functions to promote metastatic activity. Some reports suggests that AGR2 itself is secreted into the extracellular milieu (as in *Xenopus* (149)), and functions as a signalling molecule promoting the invasiveness of cancer cells (204). Others, based on the increased expression AGR2 following physiological stress (178), propose that AGR2 expression can enhance ER folding capacity, allowing increased protein production and secretion (179). While microarray studies have suggested that AGR2 signals to upregulate genes involved in proliferation, invasion and angiogenesis, factors important for metastasis and tumour progression.

Utilising the isogenic A375 cell line experimentally constitutively expressing AGR2 (Chapter 3), we aimed to interrogate whether it was possible to use proteomic methods to identify how AGR2 modified the secretome, that is the landscape of proteins secreted by the cell, resulting in a cell with increased metastatic potential. Recently, proteomics has been utilised in the detection of proteins in conditioned media from cells grown in culture, which have subsequently been used to identify changes in the secretome of the experimental cell, for example comparing inactive and active immune cells (421). Subsequently, this has been applied to oncology analysing for protein candidates which influence the metastatic characteristics of a

progressing colorectal tumour (422). It was hypothesised that an increase in AGR2 protein expression would consequently influence the maturation of client proteins in the ER, it was thought that some of these client proteins may be secretory in nature and these may be detected and quantified.

Preliminary studies required the seeding of FRT- and FRT-wtAGR2 cells at identical cell number, which were incubated for 24 hours to allow cell attachment. Next, cells were incubated with serum free media for 30 mins to remove any signalling molecules bound to the surface, followed by incubation with serum free media for 24 hours. This media, while lacking in key small molecules for cell growth and survival, would give the best opportunity to identify proteins secreted by the two cell lines. At timed intervals, samples of media were collected and snap frozen to avoid proteolytic action. Samples were concentrated using 10 kDa molecular weight cut off filters, to 1/10 of the starting volume. Subsequently, samples were boiled in SDS sample buffer and separated by SDS-PAGE. Gels were silver stained or optimal sensitivity. Observed results (Figure 7-5) suggested that proteins could be detected from the concentrated media (however it was not clear whether these were cell synthesised proteins, or as a result or cells dying through serum withdrawal, or bovine compounds retained from the FBS). wtAGR2 expressing cells suggest an increased number of protein bands identified in the first 6 hours in particular, however, it is not expected that this assay is sensitive enough to be used to quantify protein bands.

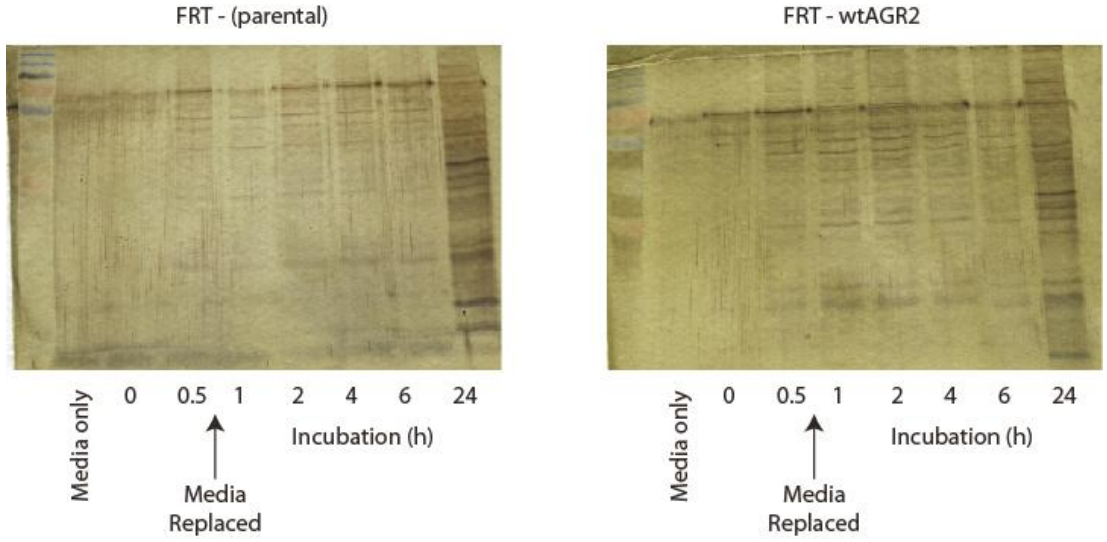


Figure 7-5 Silver staining of concentrated conditioned media from isogenic A375 cells ± AGR2 gene expression. Identical numbers of cells were plated and attached to substrate overnight. Following PBS washes, serum free media was incubated for 30 minutes before removal and replacement with fresh serum free media for the time course. At intervals 500 µL of media was removed and flash frozen for storage at -80°C, following collection all samples were concentrated using 10 kDa MWCO centrifugal filters to 50 µL, before boiling in SDS sample buffer and 25 µL loaded onto a 12% SDS-PAGE gel. Once separated gels were silver stained and imaged on a lightbox

Following these primary studies, a further experiment was prepared. Cells were similarly plated and allowed to attach to the tissue culture vessel. Following attachment, cell lines were incubated with methionine-depleted DMEM tissue culture media for 30 minutes to allow the emptying of endogenous methionine pools. Media was removed and cells were subsequently incubated with methionine-depleted media supplemented with 10% dialysed FBS and 30 μL ^{35}S-labelled methionine and incubated for 24 hours. Following ^{35}S-methionine labelling, media was removed and cells incubated with serum free media for a 24 hour time course. Samples were aspirated following 25 minutes of incubation and at time points throughout the experiment. As previously, samples were concentrated using centrifugal filters, SDS sample buffer boiled and separated by SDS PAGE. Gels were stained with Coomassie blue, and dried using a gel drier. Once dry, gels were exposed to phosphoimager screen for 16 hours before analysis on the phosphoimager (Figure 7-6). This experiment allowed labelling only of the synthesised proteins secreted from the cell and confirmed that it was possible to identify cellular proteins isolated from the media using this protocol. In addition in triplicate, 10 μL of each sample was mixed with 2 mL of Optiphase scintillation fluid and read on the scintillation counter (Figure 7-7). This indicates that synthesised proteins secreted into the media are detectable by this method, suggesting that it may be possible to quantify the protein profile of these cells allowing analysis of the effect of AGR2 expression on secreted proteins.

It is proposed that further quantitative mass spectrometric analysis, such as SILAC or ITAQ, may highlight important changes in the abundance of proteins secreted from the cell and whether these have a bearing on the metastatic potential of the cell as a result of AGR2 gene expression.

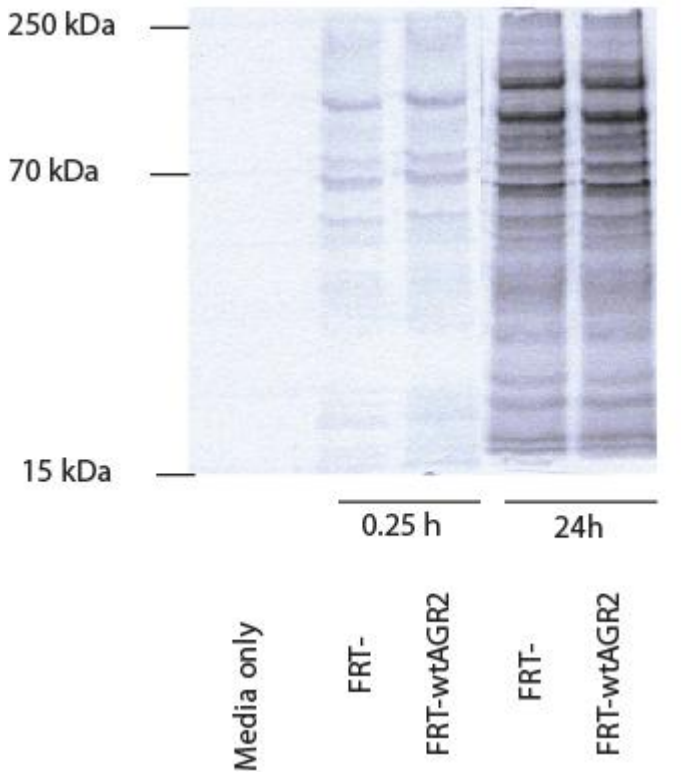


Figure 7-6 Phosphoimager screen of FRT- and FRT-wtAGR2 cells incubated with ^{35}S-methionine supplemented DMEM tissue culture media. Following labelling cells were washed and incubated with serum free DMEM for 24 hour time course. Samples presented here were prior to incubation (media only) and following 15 minutes and 24 hours of incubation with serum free media. Media was concentrated to 1/10th of starting volume using centrifugal filters prior to SDS-PAGE and the gel dried. To detect dried gels were exposed to the phosphimager screen for 16 hours before analysis on the phosphoimager.

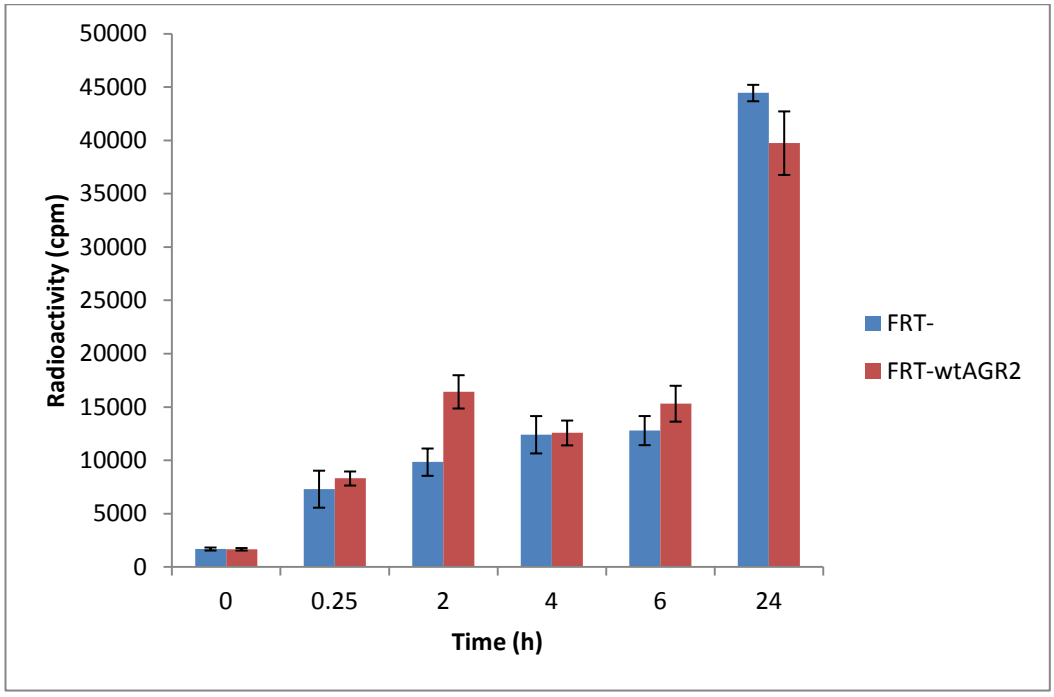


Figure 7-7 Scintillation count analysis of FRT- and FRT-wtAGR2 cells incubated with ^{35}S-methionine supplemented DMEM tissue culture media. Following labelling cells were washed and incubated with serum free DMEM for 24 hour time course. In triplicate 10 µL of each sample was mixed with 2 mL of Optiphase scintillation fluid prior to quantitation on the scintillation counter.

Chapter 8: Appendices

Compact disc attached to back cover

8.1 Appendix 1 – Illumina HT-12 microarray data

Microsoft Excel file: Appendix 1 Microarray Table.xlsx

8.2 Appendix 2 – SILAC protein expression data

Microsoft Excel file: Appendix 2 Supplementary SILAC table.xlsx

Chapter 9: References

s(1) World Health Organisation. WHO 2013 Mar 18.

(2) Mistry M, Parkin DM, Ahmad AS, Sasieni P. Cancer incidence in the United Kingdom: projections to the year 2030. Br J Cancer 2011 Nov 22;105(11):1795-803.

(3) Jones PA, Baylin SB. The epigenomics of cancer. Cell 2007 Feb 23;128(4):683-92.

(4) Loeb LA, Loeb KR, Anderson JP. Multiple mutations and cancer. Proc Natl Acad Sci U S A 2003 Feb 4;100(3):776-81.

(5) Vogelstein B, Kinzler KW. Cancer genes and the pathways they control. Nat Med 2004 Aug;10(8):789-99.

(6) Hanahan D, Weinberg RA. The hallmarks of cancer. Cell 2000 Jan 7;100(1):57-70.

(7) Hahn WC, Counter CM, Lundberg AS, Beijersbergen RL, Brooks MW, Weinberg RA. Creation of human tumour cells with defined genetic elements. Nature 1999 Jul 29;400(6743):464-8.

(8) Fearon ER, Vogelstein B. A genetic model for colorectal tumorigenesis. Cell 1990 Jun 1;61(5):759-67.

(9) Holliday R. The inheritance of epigenetic defects. Science 1987 Oct 9;238(4824):163-70.

(10) Lund AH, van LM. Epigenetics and cancer. Genes Dev 2004 Oct 1;18(19):2315-35.

(11) Fraga MF, Ballestar E, Villar-Garea A, Boix-Chornet M, Espada J, Schotta G, et al. Loss of acetylation at Lys16 and trimethylation at Lys20 of histone H4 is a common hallmark of human cancer. Nat Genet 2005 Apr;37(4):391-400.

(12) Feinberg AP, Vogelstein B. Hypomethylation distinguishes genes of some human cancers from their normal counterparts. Nature 1983 Jan 6;301(5895):89-92.

(13) Herman JG. Hypermethylation of tumor suppressor genes in cancer. Semin Cancer Biol 1999 Oct;9(5):359-67.

(14) Warburg O. [Origin of cancer cells]. Oncologia 1956;9(2):75-83.

(15) Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell 2011 Mar 4;144(5):646-74.

(16) Allen M, Louise JJ. Jekyll and Hyde: the role of the microenvironment on the progression of cancer. J Pathol 2011 Jan;223(2):162-76.

(17) Begg AC, Stewart FA, Vens C. Strategies to improve radiotherapy with targeted drugs. Nat Rev Cancer 2011 Apr;11(4):239-53.

(18) Lopes NM, Adams EG, Pitts TW, Bhuyan BK. Cell kill kinetics and cell cycle effects of taxol on human and hamster ovarian cell lines. Cancer Chemother Pharmacol 1993;32(3):235-42.

(19) Rowinsky EK, Cazenave LA, Donehower RC. Taxol: a novel investigational antimicrotubule agent. J Natl Cancer Inst 1990 Aug 1;82(15):1247-59.

(20) Bharadwaj R, Yu H. The spindle checkpoint, aneuploidy, and cancer. Oncogene 2004 Mar 15;23(11):2016-27.

(21) Siddik ZH. Cisplatin: mode of cytotoxic action and molecular basis of resistance. Oncogene 2003 Oct 20;22(47):7265-79.

(22) Bonvini P, Zorzi E, Basso G, Rosolen A. Bortezomib-mediated 26S proteasome inhibition causes cell-cycle arrest and induces apoptosis in CD-30+ anaplastic large cell lymphoma. Leukemia 2007 Apr;21(4):838-42.

(23) Futreal PA, Coin L, Marshall M, Down T, Hubbard T, Wooster R, et al. A census of human cancer genes. Nat Rev Cancer 2004 Mar;4(3):177-83.

(24) Weinstein IB, Joe AK. Mechanisms of disease: Oncogene addiction--a rationale for molecular targeting in cancer therapy. Nat Clin Pract Oncol 2006 Aug;3(8):448-57.

(25) Shaul Y. c-Abl: activation and nuclear targets. Cell Death Differ 2000 Jan;7(1):10-6.

(26) Druker BJ. Inhibition of the Bcr-Abl tyrosine kinase as a therapeutic strategy for CML. Oncogene 2002 Dec 9;21(56):8541-6.

(27) Rusch V, Klimstra D, Venkatraman E, Pisters PW, Langenfeld J, Dmitrovsky E. Overexpression of the epidermal growth factor receptor and its ligand transforming growth factor alpha is frequent in resectable non-small cell lung cancer but does not predict tumor progression. Clin Cancer Res 1997 Apr;3(4):515-22.

(28) Davies H, Bignell GR, Cox C, Stephens P, Edkins S, Clegg S, et al. Mutations of the BRAF gene in human cancer. Nature 2002 Jun 27;417(6892):949-54.

(29) Cho HS, Mason K, Ramyar KX, Stanley AM, Gabelli SB, Denney DW, Jr., et al. Structure of the extracellular region of HER2 alone and in complex with the Herceptin Fab. Nature 2003 Feb 13;421(6924):756-60.

(30) Barinaga M. From bench top to bedside. Science 1997 Nov 7;278(5340):1036-9.

(31) Gottesman MM. Mechanisms of cancer drug resistance. Annu Rev Med 2002;53:615-27.

(32) Goh KI, Choi IG. Exploring the human diseasome: the human disease network. Brief Funct Genomics 2012 Nov;11(6):533-42.

(33) Coulombe B, Jeronimo C, Langelier MF, Cojocaru M, Bergeron D. Interaction networks of the molecular machines that decode, replicate, and maintain the integrity of the human genome. Mol Cell Proteomics 2004 Sep;3(9):851-6.

(34) Sanchez C, Lachaize C, Janody F, Bellon B, Roder L, Euzenat J, et al. Grasping at molecular interactions and genetic networks in Drosophila melanogaster using FlyNets, an Internet database. Nucleic Acids Res 1999 Jan 1;27(1):89-94.

(35) Joerger AC, Fersht AR. Structural biology of the tumor suppressor p53. Annu Rev Biochem 2008;77:557-82.

(36) Maslon MM, Hupp TR. Drug discovery and mutant p53. Trends Cell Biol 2010 Sep;20(9):542-55.

(37) Olivier M, Eeles R, Hollstein M, Khan MA, Harris CC, Hainaut P. The IARC TP53 database: new online mutation analysis and recommendations to users. Hum Mutat 2002 Jun;19(6):607-14.

(38) Lambert JM, Gorzov P, Veprintsev DB, Soderqvist M, Segerback D, Bergman J, et al. PRIMA-1 reactivates mutant p53 by covalent binding to the core domain. Cancer Cell 2009 May 5;15(5):376-88.

(39) Jolly C, Morimoto RI. Role of the heat shock response and molecular chaperones in oncogenesis and cell death. J Natl Cancer Inst 2000 Oct 4;92(19):1564-72.

(40) Whitesell L, Lindquist SL. HSP90 and the chaperoning of cancer. Nat Rev Cancer 2005 Oct;5(10):761-72.

(41) Powers ET, Morimoto RI, Dillin A, Kelly JW, Balch WE. Biological and chemical approaches to diseases of proteostasis deficiency. Annu Rev Biochem 2009;78:959-91.

(42) Calderwood SK, Khaleque MA, Sawyer DB, Ciocca DR. Heat shock proteins in cancer: chaperones of tumorigenesis. Trends Biochem Sci 2006 Mar;31(3):164-72.

(43) Whitesell L, Mimnaugh EG, De CB, Myers CE, Neckers LM. Inhibition of heat shock protein HSP90-pp60v-src heteroprotein complex formation by

benzoquinone ansamycins: essential role for stress proteins in oncogenic transformation. Proc Natl Acad Sci U S A 1994 Aug 30;91(18):8324-8.

(44) Georgakis GV, Li Y, Rassidakis GZ, Martinez-Valdez H, Medeiros LJ, Younes A. Inhibition of heat shock protein 90 function by 17-allylamino-17-demethoxy-geldanamycin in Hodgkin's lymphoma cells down-regulates Akt kinase, dephosphorylates extracellular signal-regulated kinase, and induces cell cycle arrest and cell death. Clin Cancer Res 2006 Jan 15;12(2):584-90.

(45) Solar P, Horvath V, Kleban J, Koval' J, Solarova Z, Kozubik A, et al. Hsp90 inhibitor geldanamycin increases the sensitivity of resistant ovarian adenocarcinoma cell line A2780cis to cisplatin. Neoplasma 2007;54(2):127-30.

(46) Sugimoto K, Sasaki M, Isobe Y, Tsutsui M, Suto H, Ando J, et al. Hsp90-inhibitor geldanamycin abrogates G2 arrest in p53-negative leukemia cell lines through the depletion of Chk1. Oncogene 2008 May 15;27(22):3091-101.

(47) Reich NC, Levine AJ. Specific interaction of the SV40 T antigen-cellular p53 protein complex with SV40 DNA. Virology 1982 Feb;117(1):286-90.

(48) Hollstein M, Rice K, Greenblatt MS, Soussi T, Fuchs R, Sorlie T, et al. Database of p53 gene somatic mutations in human tumors and cell lines. Nucleic Acids Res 1994 Sep;22(17):3551-5.

(49) Donehower LA, Harvey M, Slagle BL, McArthur MJ, Montgomery CA, Jr., Butel JS, et al. Mice deficient for p53 are developmentally normal but susceptible to spontaneous tumours. Nature 1992 Mar 19;356(6366):215-21.

(50) Lutzker SG, Levine AJ. A functionally inactive p53 protein in teratocarcinoma cells is activated by either DNA damage or cellular differentiation. Nat Med 1996 Jul;2(7):804-10.

(51) Ostermeyer AG, Runko E, Winkfield B, Ahn B, Moll UM. Cytoplasmically sequestered wild-type p53 protein in neuroblastoma is relocated to the nucleus by a C-terminal peptide. Proc Natl Acad Sci U S A 1996 Dec 24;93(26):15190-4.

(52) Moll UM, Wolff S, Speidel D, Deppert W. Transcription-independent pro-apoptotic functions of p53. Curr Opin Cell Biol 2005 Dec;17(6):631-6.

(53) Moll UM, Riou G, Levine AJ. Two distinct mechanisms alter p53 in breast cancer: mutation and nuclear exclusion. Proc Natl Acad Sci U S A 1992 Aug 1;89(15):7262-6.

(54) Moll UM, Ostermeyer AG, Haladay R, Winkfield B, Frazier M, Zambetti G. Cytoplasmic sequestration of wild-type p53 protein impairs the G1 checkpoint after DNA damage. Mol Cell Biol 1996 Mar;16(3):1126-37.

(55) Dornan D, Shimizu H, Burch L, Smith AJ, Hupp TR. The proline repeat domain of p53 binds directly to the transcriptional coactivator p300 and allosterically controls DNA-dependent acetylation of p53. Mol Cell Biol 2003 Dec;23(23):8846-61.

(56) Kussie PH, Gorina S, Marechal V, Elenbaas B, Moreau J, Levine AJ, et al. Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor transactivation domain. Science 1996 Nov 8;274(5289):948-53.

(57) Soussi T, May P. Structural aspects of the p53 protein in relation to gene evolution: a second look. J Mol Biol 1996 Aug 2;260(5):623-37.

(58) Soussi T, Caron de FC, May P. Structural aspects of the p53 protein in relation to gene evolution. Oncogene 1990 Jul;5(7):945-52.

(59) Emamzadah S, Tropia L, Halazonetis TD. Crystal structure of a multidomain human p53 tetramer bound to the natural CDKN1A (p21) p53-response element. Mol Cancer Res 2011 Nov;9(11):1493-9.

(60) Malecka KA, Ho WC, Marmorstein R. Crystal structure of a p53 core tetramer bound to DNA. Oncogene 2009 Jan 22;28(3):325-33.

(61) Ma B, Pan Y, Zheng J, Levine AJ, Nussinov R. Sequence analysis of p53 response-elements suggests multiple binding modes of the p53 tetramer to DNA targets. Nucleic Acids Res 2007;35(9):2986-3001.

(62) Friedman PN, Chen X, Bargonetti J, Prives C. The p53 protein is an unusually shaped tetramer that binds directly to DNA. Proc Natl Acad Sci U S A 1993 Apr 15;90(8):3319-23.

(63) McLure KG, Lee PW. How p53 binds DNA as a tetramer. EMBO J 1998 Jun 15;17(12):3342-50.

(64) Gotz C, Scholtes P, Prowald A, Schuster N, Nastainczyk W, Montenarh M. Protein kinase CK2 interacts with a multi-protein binding domain of p53. Mol Cell Biochem 1999 Jan;191(1-2):111-20.

(65) van DJ, Fernandez-Fernandez MR, Veprintsev DB, Fersht AR. Modulation of the oligomerization state of p53 by differential binding of proteins of the S100 family to p53 monomers and tetramers. J Biol Chem 2009 May 15;284(20):13804-11.

(66) Fernandez-Fernandez MR, Rutherford TJ, Fersht AR. Members of the S100 family bind p53 in two distinct ways. Protein Sci 2008 Oct;17(10):1663-70.

(67) Rustandi RR, Baldisseri DM, Weber DJ. Structure of the negative regulatory domain of p53 bound to S100B(betabeta). Nat Struct Biol 2000 Jul;7(7):570-4.

(68) Mujtaba S, He Y, Zeng L, Yan S, Plotnikova O, Sachchidanand, et al. Structural mechanism of the bromodomain of the coactivator CBP in p53 transcriptional activation. Mol Cell 2004 Jan 30;13(2):251-63.

(69) Natan E, Baloglu C, Pagel K, Freund SM, Morgner N, Robinson CV, et al. Interaction of the p53 DNA-binding domain with its n-terminal extension modulates the stability of the p53 tetramer. J Mol Biol 2011 Jun 10;409(3):358-68.

(70) Wells M, Tidow H, Rutherford TJ, Markwick P, Jensen MR, Mylonas E, et al. Structure of tumor suppressor p53 and its intrinsically disordered N-terminal transactivation domain. Proc Natl Acad Sci U S A 2008 Apr 15;105(15):5762-7.

(71) Blaydes JP, Luciani MG, Pospisilova S, Ball HM, Vojtesek B, Hupp TR. Stoichiometric phosphorylation of human p53 at Ser315 stimulates p53-dependent transcription. J Biol Chem 2001 Feb 16;276(7):4699-708.

(72) Hupp TR, Lane DP. Regulation of the cryptic sequence-specific DNA-binding function of p53 by protein kinases. Cold Spring Harb Symp Quant Biol 1994;59:195-206.

(73) Appella E, Anderson CW. Post-translational modifications and activation of p53 by genotoxic stresses. Eur J Biochem 2001 May;268(10):2764-72.

(74) Cho Y, Gorina S, Jeffrey PD, Pavletich NP. Crystal structure of a p53 tumor suppressor-DNA complex: understanding tumorigenic mutations. Science 1994 Jul 15;265(5170):346-55.

(75) Tsai CJ, Ma B, Nussinov R. Protein-protein interaction networks: how can a hub protein bind so many different partners? Trends Biochem Sci 2009 Dec;34(12):594-600.

(76) Shimizu S, Kanaseki T, Mizushima N, Mizuta T, Arakawa-Kobayashi S, Thompson CB, et al. Role of Bcl-2 family proteins in a non-apoptotic programmed cell death dependent on autophagy genes. Nat Cell Biol 2004 Dec;6(12):1221-8.

(77) Levine AJ. p53, the cellular gatekeeper for growth and division. Cell 1997 Feb 7;88(3):323-31.

(78) Lane DP. Cancer. p53, guardian of the genome. Nature 1992 Jul 2;358(6381):15-6.

(79) Lakin ND, Jackson SP. Regulation of p53 in response to DNA damage. Oncogene 1999 Dec 13;18(53):7644-55.

(80) Momand J, Wu HH, Dasgupta G. MDM2--master regulator of the p53 tumor suppressor protein. Gene 2000 Jan 25;242(1-2):15-29.

(81) Maltzman W, Czyzyk L. UV irradiation stimulates levels of p53 cellular tumor antigen in nontransformed mouse cells. Mol Cell Biol 1984 Sep;4(9):1689-94.

(82) Kastan MB, Onyekwere O, Sidransky D, Vogelstein B, Craig RW. Participation of p53 protein in the cellular response to DNA damage. Cancer Res 1991 Dec 1;51(23 Pt 1):6304-11.

(83) Davison TS, Yin P, Nie E, Kay C, Arrowsmith CH. Characterization of the oligomerization defects of two p53 mutants found in families with Li-Fraumeni and Li-Fraumeni-like syndrome. Oncogene 1998 Aug 6;17(5):651-6.

(84) El-Deiry WS, Tokino T, Velculescu VE, Levy DB, Parsons R, Trent JM, et al. WAF1, a potential mediator of p53 tumor suppression. Cell 1993 Nov 19;75(4):817-25.

(85) Ko LJ, Prives C. p53: puzzle and paradigm. Genes Dev 1996 May 1;10(9):1054-72.

(86) Bode AM, Dong Z. Post-translational modification of p53 in tumorigenesis. Nat Rev Cancer 2004 Oct;4(10):793-805.

(87) Nakano K, Vousden KH. PUMA, a novel proapoptotic gene, is induced by p53. Mol Cell 2001 Mar;7(3):683-94.

(88) Oda E, Ohki R, Murasawa H, Nemoto J, Shibue T, Yamashita T, et al. Noxa, a BH3-only member of the Bcl-2 family and candidate mediator of p53-induced apoptosis. Science 2000 May 12;288(5468):1053-8.

(89) Sax JK, Fei P, Murphy ME, Bernhard E, Korsmeyer SJ, El-Deiry WS. BID regulation by p53 contributes to chemosensitivity. Nat Cell Biol 2002 Nov;4(11):842-9.

(90) Chipuk JE, Kuwana T, Bouchier-Hayes L, Droin NM, Newmeyer DD, Schuler M, et al. Direct activation of Bax by p53 mediates mitochondrial membrane permeabilization and apoptosis. Science 2004 Feb 13;303(5660):1010-4.

(91) Vousden KH, Prives C. Blinded by the Light: The Growing Complexity of p53. Cell 2009 May 1;137(3):413-31.

(92) Laptenko O, Prives C. Transcriptional regulation by p53: one protein, many possibilities. Cell Death Differ 2006 Jun;13(6):951-61.

(93) Youle RJ, Strasser A. The BCL-2 protein family: opposing activities that mediate cell death. Nat Rev Mol Cell Biol 2008 Jan;9(1):47-59.

(94) Li J, Yuan J. Caspases in apoptosis and beyond. Oncogene 2008 Oct 20;27(48):6194-206.

(95) Wagner AJ, Kokontis JM, Hay N. Myc-mediated apoptosis requires wild-type p53 in a manner independent of cell cycle arrest and the ability of p53 to induce p21waf1/cip1. Genes Dev 1994 Dec 1;8(23):2817-30.

(96) Caelles C, Helmberg A, Karin M. p53-dependent apoptosis in the absence of transcriptional activation of p53-target genes. Nature 1994 Jul 21;370(6486):220-3.

(97) Haupt Y, Rowan S, Shaulian E, Vousden KH, Oren M. Induction of apoptosis in HeLa cells by trans-activation-deficient p53. Genes Dev 1995 Sep 1;9(17):2170-83.

(98) Chipuk JE, Green DR. p53's believe it or not: lessons on transcription-independent death. J Clin Immunol 2003 Sep;23(5):355-61.

(99) Speidel D. Transcription-independent p53 apoptosis: an alternative route to death. Trends Cell Biol 2010 Jan;20(1):14-24.

(100) Jiang P, Du W, Heese K, Wu M. The Bad guy cooperates with good cop p53: Bad is transcriptionally up-regulated by p53 and forms a Bad/p53 complex at the mitochondria to induce apoptosis. Mol Cell Biol 2006 Dec;26(23):9071-82.

(101) Yu J, Zhang L. No PUMA, no death: implications for p53-dependent apoptosis. Cancer Cell 2003 Oct;4(4):248-9.

(102) Chan TA, Hwang PM, Hermeking H, Kinzler KW, Vogelstein B. Cooperative effects of genes controlling the G(2)/M checkpoint. Genes Dev 2000 Jul 1;14(13):1584-8.

(103) Jackson JG, Pereira-Smith OM. p53 is preferentially recruited to the promoters of growth arrest genes p21 and GADD45 during replicative senescence of normal human fibroblasts. Cancer Res 2006 Sep 1;66(17):8356-60.

(104) Bouvet M, Ellis LM, Nishizaki M, Fujiwara T, Liu W, Bucana CD, et al. Adenovirus-mediated wild-type p53 gene transfer down-regulates vascular endothelial growth factor expression and inhibits angiogenesis in human colon cancer. Cancer Res 1998 Jun 1;58(11):2288-92.

(105) Kwak C, Jin RJ, Lee C, Park MS, Lee SE. Thrombospondin-1, vascular endothelial growth factor expression and their relationship with p53 status in prostate cancer and benign prostatic hyperplasia. BJU Int 2002 Feb;89(3):303-9.

(106) Ravi R, Mookerjee B, Bhujwalla ZM, Sutter CH, Artemov D, Zeng Q, et al. Regulation of tumor angiogenesis by p53-induced degradation of hypoxia-inducible factor 1alpha. Genes Dev 2000 Jan 1;14(1):34-44.

(107) El-Deiry WS, Harper JW, O'Connor PM, Velculescu VE, Canman CE, Jackman J, et al. WAF1/CIP1 is induced in p53-mediated G1 arrest and apoptosis. Cancer Res 1994 Mar 1;54(5):1169-74.

(108) Stewart ZA, Leach SD, Pietenpol JA. p21(Waf1/Cip1) inhibition of cyclin E/Cdk2 activity prevents endoreduplication after mitotic spindle disruption. Mol Cell Biol 1999 Jan;19(1):205-15.

(109) de La FC, Santiago F, Chong SY, Deng L, Mayhood T, Fu P, et al. Overexpression of p21(waf1) in human T-cell lymphotropic virus type 1-infected cells and its association with cyclin A/cdk2. J Virol 2000 Aug;74(16):7270-83.

(110) El-Deiry WS, Tokino T, Waldman T, Oliner JD, Velculescu VE, Burrell M, et al. Topological control of p21WAF1/CIP1 expression in normal and neoplastic tissues. Cancer Res 1995 Jul 1;55(13):2910-9.

(111) Ocker M, Schneider-Stock R. Histone deacetylase inhibitors: signalling towards p21cip1/waf1. Int J Biochem Cell Biol 2007;39(7-8):1367-74.

(112) Bornstein G, Bloom J, Sitry-Shevah D, Nakayama K, Pagano M, Hershko A. Role of the SCFSkp2 ubiquitin ligase in the degradation of p21Cip1 in S phase. J Biol Chem 2003 Jul 11;278(28):25752-7.

(113) Bendjennat M, Boulaire J, Jascur T, Brickner H, Barbier V, Sarasin A, et al. UV irradiation triggers ubiquitin-dependent degradation of p21(WAF1) to promote DNA repair. Cell 2003 Sep 5;114(5):599-610.

(114) Pang LY, Scott M, Hayward RL, Mohammed H, Whitelaw CB, Smith GC, et al. p21(WAF1) is component of a positive feedback loop that maintains the p53 transcriptional program. Cell Cycle 2011 Mar 15;10(6):932-50.

(115) Rodrigues NR, Rowan A, Smith ME, Kerr IB, Bodmer WF, Gannon JV, et al. p53 mutations in colorectal cancer. Proc Natl Acad Sci U S A 1990 Oct;87(19):7555-9.

(116) Milner J, Medcalf EA. Cotranslation of activated mutant p53 with wild type drives the wild-type p53 protein into the mutant conformation. Cell 1991 May 31;65(5):765-74.

(117) Kern SE, Kinzler KW, Bruskin A, Jarosz D, Friedman P, Prives C, et al. Identification of p53 as a sequence-specific DNA-binding protein. Science 1991 Jun 21;252(5013):1708-11.

(118) Dittmer D, Pati S, Zambetti G, Chu S, Teresky AK, Moore M, et al. Gain of function mutations in p53. Nat Genet 1993 May;4(1):42-6.

(119) Wolf D, Harris N, Rotter V. Reconstitution of p53 expression in a nonproducer Ab-MuLV-transformed cell line by transfection of a functional p53 gene. Cell 1984 Aug;38(1):119-26.

(120) Baker SJ, Preisinger AC, Jessup JM, Paraskeva C, Markowitz S, Willson JK, et al. p53 gene mutations occur in combination with 17p allelic deletions as late events in colorectal tumorigenesis. Cancer Res 1990 Dec 1;50(23):7717-22.

(121) Ohgaki H, Dessen P, Jourde B, Horstmann S, Nishikawa T, Di Patre PL, et al. Genetic pathways to glioblastoma: a population-based study. Cancer Res 2004 Oct 1;64(19):6892-9.

(122) Vogelstein B, Lane D, Levine AJ. Surfing the p53 network. Nature 2000 Nov 16;408(6810):307-10.

(123) Scheffner M, Nuber U, Huibregtse JM. Protein ubiquitination involving an E1-E2-E3 enzyme ubiquitin thioester cascade. Nature 1995 Jan 5;373(6509):81-3.

(124) Roth J, Dobbelstein M, Freedman DA, Shenk T, Levine AJ. Nucleo-cytoplasmic shuttling of the hdm2 oncoprotein regulates the levels of the p53 protein via a pathway used by the human immunodeficiency virus rev protein. EMBO J 1998 Jan 15;17(2):554-64.

(125) Tao W, Levine AJ. Nucleocytoplasmic shuttling of oncoprotein Hdm2 is required for Hdm2-mediated degradation of p53. Proc Natl Acad Sci U S A 1999 Mar 16;96(6):3077-80.

(126) O'Keefe K, Li H, Zhang Y. Nucleocytoplasmic shuttling of p53 is essential for MDM2-mediated cytoplasmic degradation but not ubiquitination. Mol Cell Biol 2003 Sep;23(18):6396-405.

(127) Jones SN, Roe AE, Donehower LA, Bradley A. Rescue of embryonic lethality in Mdm2-deficient mice by absence of p53. Nature 1995 Nov 9;378(6553):206-8.

(128) Chen J, Wu X, Lin J, Levine AJ. mdm-2 inhibits the G1 arrest and apoptosis functions of the p53 tumor suppressor protein. Mol Cell Biol 1996 May;16(5):2445-52.

(129) Eymin B, Gazzeri S, Brambilla C, Brambilla E. Mdm2 overexpression and p14(ARF) inactivation are two mutually exclusive events in primary human lung tumors. Oncogene 2002 Apr 18;21(17):2750-61.

(130) Leite KR, Franco MF, Srougi M, Nesrallah LJ, Nesrallah A, Bevilacqua RG, et al. Abnormal expression of MDM2 in prostate carcinoma. Mod Pathol 2001 May;14(5):428-36.

(131) Polsky D, Bastian BC, Hazan C, Melzer K, Pack J, Houghton A, et al. HDM2 protein overexpression, but not gene amplification, is related to tumorigenesis of cutaneous melanoma. Cancer Res 2001 Oct 15;61(20):7642-6.

(132) Foster BA, Coffey HA, Morin MJ, Rastinejad F. Pharmacological rescue of mutant p53 conformation and function. Science 1999 Dec 24;286(5449):2507-10.

(133) Rippin TM, Bykov VJ, Freund SM, Selivanova G, Wiman KG, Fersht AR. Characterization of the p53-rescue drug CP-31398 in vitro and in living cells. Oncogene 2002 Mar 28;21(14):2119-29.

(134) Takimoto R, Wang W, Dicker DT, Rastinejad F, Lyssikatos J, El-Deiry WS. The mutant p53-conformation modifying drug, CP-31398, can induce apoptosis of human cancer cells and can stabilize wild-type p53 protein. Cancer Biol Ther 2002 Jan;1(1):47-55.

(135) Basse N, Kaar JL, Settanni G, Joerger AC, Rutherford TJ, Fersht AR. Toward the rational design of p53-stabilizing drugs: probing the surface of the oncogenic Y220C mutant. Chem Biol 2010 Jan 29;17(1):46-56.

(136) Boeckler FM, Joerger AC, Jaggi G, Rutherford TJ, Veprintsev DB, Fersht AR. Targeted rescue of a destabilized mutant of p53 by an in silico screened drug. Proc Natl Acad Sci U S A 2008 Jul 29;105(30):10360-5.

(137) Woods DB, Vousden KH. Regulation of p53 function. Exp Cell Res 2001 Mar 10;264(1):56-66.

(138) Candeias MM, Malbert-Colas L, Powell DJ, Daskalogianni C, Maslon MM, Naski N, et al. P53 mRNA controls p53 activity by managing Mdm2 functions. Nat Cell Biol 2008 Sep;10(9):1098-105.

(139) Vassilev LT. MDM2 inhibitors for cancer therapy. Trends Mol Med 2007 Jan;13(1):23-31.

(140) Chene P. Inhibiting the p53-MDM2 interaction: an important target for cancer therapy. Nat Rev Cancer 2003 Feb;3(2):102-9.

(141) Vassilev LT. Small-molecule antagonists of p53-MDM2 binding: research tools and potential therapeutics. Cell Cycle 2004 Apr;3(4):419-21.

(142) Tovar C, Rosinski J, Filipovic Z, Higgins B, Kolinsky K, Hilton H, et al. Small-molecule MDM2 antagonists reveal aberrant p53 signaling in cancer: implications for therapy. Proc Natl Acad Sci U S A 2006 Feb 7;103(6):1888-93.

(143) Nieves-Neira W, Rivera MI, Kohlhagen G, Hursey ML, Pourquier P, Sausville EA, et al. DNA protein cross-links produced by NSC 652287, a novel thiophene derivative active against human renal cancer cells. Mol Pharmacol 1999 Sep;56(3):478-84.

(144) Rivera MI, Stinson SF, Vistica DT, Jorden JL, Kenney S, Sausville EA. Selective toxicity of the tricyclic thiophene NSC 652287 in renal carcinoma cell lines: differential accumulation and metabolism. Biochem Pharmacol 1999 Jun 1;57(11):1283-95.

(145) Issaeva N, Bozko P, Enge M, Protopopova M, Verhoef LG, Masucci M, et al. Small molecule RITA binds to p53, blocks p53-HDM-2 interaction and activates p53 function in tumors. Nat Med 2004 Dec;10(12):1321-8.

(146) Spinnler C, Hedstrom E, Li H, de LJ, Nikulenkov F, Teunisse AF, et al. Abrogation of Wip1 expression by RITA-activated p53 potentiates apoptosis induction via activation of ATM and inhibition of HdmX. Cell Death Differ 2011 Nov;18(11):1736-45.

(147) Wang SP, Wang WL, Chang YL, Wu CT, Chao YC, Kao SH, et al. p53 controls cancer cell invasion by inducing the MDM2-mediated degradation of Slug. Nat Cell Biol 2009 Jun;11(6):694-704.

(148) Sive HL, Hattori K, Weintraub H. Progressive determination during formation of the anteroposterior axis in Xenopus laevis. Cell 1989 Jul 14;58(1):171-80.

(149) Aberger F, Weidinger G, Grunz H, Richter K. Anterior specification of embryonic ectoderm: the role of the Xenopus cement gland-specific gene XAG-2. Mech Dev 1998 Mar;72(1-2):115-30.

(150) Pohler E, Craig AL, Cotton J, Lawrie L, Dillon JF, Ross P, et al. The Barrett's antigen anterior gradient-2 silences the p53 transcriptional response to DNA damage. Mol Cell Proteomics 2004 Jun;3(6):534-47.

(151) Bradley L, Wainstock D, Sive H. Positive and negative signals modulate formation of the Xenopus cement gland. Development 1996 Sep;122(9):2739-50.

(152) Sive H, Bradley L. A sticky problem: the Xenopus cement gland as a paradigm for anteroposterior patterning. Dev Dyn 1996 Mar;205(3):265-80.

(153) Novoselov VV, Alexandrova EM, Ermakova GV, Zaraisky AG. Expression zones of three novel genes abut the developing anterior neural plate of Xenopus embryo. Gene Expr Patterns 2003 May;3(2):225-30.

(154) Kumar A, Godwin JW, Gates PB, Garza-Garcia AA, Brockes JP. Molecular basis for the nerve dependence of limb regeneration in an adult vertebrate. Science 2007 Nov 2;318(5851):772-7.

(155) da Silva SM, Gates PB, Brockes JP. The newt ortholog of CD59 is implicated in proximodistal identity during amphibian limb regeneration. Dev Cell 2002 Oct;3(4):547-55.

(156) Wood AJ, Lo TW, Zeitler B, Pickle CS, Ralston EJ, Lee AH, et al. Targeted genome editing across species using ZFNs and TALENs. Science 2011 Jul 15;333(6040):307.

(157) Ivanova AS, Tereshina MB, Ermakova GV, Belousov VV, Zaraisky AG. Agr genes, missing in amniotes, are involved in the body appendages regeneration in frog tadpoles. Sci Rep 2013 Feb 15;3:1279.

(158) Shih LJ, Lu YF, Chen YH, Lin CC, Chen JA, Hwang SP. Characterization of the agr2 gene, a homologue of X. laevis anterior gradient 2, from the zebrafish, Danio rerio. Gene Expr Patterns 2007 Feb;7(4):452-60.

(159) Chen YC, Lu YF, Li IC, Hwang SP. Zebrafish Agr2 is required for terminal differentiation of intestinal goblet cells. PLoS One 2012;7(4):e34408.

(160) Morrison RN, Cooper GA, Koop BF, Rise ML, Bridle AR, Adams MB, et al. Transcriptome profiling the gills of amoebic gill disease (AGD)-affected Atlantic salmon (Salmo salar L.): a role for tumor suppressor p53 in AGD pathogenesis? Physiol Genomics 2006 Jun 16;26(1):15-34.

(161) Xia JH, Jiang J, Shi YH, Gui JF. Predominant expression and cellular distribution of fish Agr2 in renal collecting system. Comp Biochem Physiol B Biochem Mol Biol 2009 Apr;152(4):397-404.

(162) Komiya T, Tanigawa Y, Hirohashi S. Cloning of the gene gob-4, which is expressed in intestinal goblet cells in mice. Biochim Biophys Acta 1999 Mar 19;1444(3):434-8.

(163) Cheng H, Leblond CP. Origin, differentiation and renewal of the four main epithelial cell types in the mouse small intestine. V. Unitarian Theory of the origin of the four epithelial cell types. Am J Anat 1974 Dec;141(4):537-61.

(164) Wang Z, Hao Y, Lowe AW. The adenocarcinoma-associated antigen, AGR2, promotes tumor growth, cell migration, and cellular transformation. Cancer Res 2008 Jan 15;68(2):492-7.

(165) Park SW, Zhen G, Verhaeghe C, Nakagami Y, Nguyenvu LT, Barczak AJ, et al. The protein disulfide isomerase AGR2 is essential for production of intestinal mucus. Proc Natl Acad Sci U S A 2009 Apr 28;106(17):6950-5.

(166) Zhao F, Edwards R, Dizon D, Afrasiabi K, Mastroianni JR, Geyfman M, et al. Disruption of Paneth and goblet cell homeostasis and increased endoplasmic reticulum stress in Agr2-/- mice. Dev Biol 2010 Feb 15;338(2):270-9.

(167) Keshav S, Lawson L, Chung LP, Stein M, Perry VH, Gordon S. Tumor necrosis factor mRNA localized to Paneth cells of normal murine intestinal epithelium by in situ hybridization. J Exp Med 1990 Jan 1;171(1):327-32.

(168) Ganz T. Paneth cells--guardians of the gut cell hatchery. Nat Immunol 2000 Aug;1(2):99-100.

(169) Chevet E, Fessart D, Delom F, Mulot A, Vojtesek B, Hrstka R, et al. Emerging roles for the pro-oncogenic anterior gradient-2 in cancer development. Oncogene 2012 Sep 3.

(170) Petek E, Windpassinger C, Egger H, Kroisel PM, Wagner K. Localization of the human anterior gradient-2 gene (AGR2) to chromosome band 7p21.3 by radiation hybrid mapping and fluorescencein situ hybridisation. Cytogenet Cell Genet 2000;89(3-4):141-2.

(171) Zheng W, Rosenstiel P, Huse K, Sina C, Valentonyte R, Mah N, et al. Evaluation of AGR2 and AGR3 as candidate genes for inflammatory bowel disease. Genes Immun 2006 Jan;7(1):11-8.

(172) Fletcher GC, Patel S, Tyson K, Adam PJ, Schenker M, Loader JA, et al. hAG-2 and hAG-3, human homologues of genes involved in differentiation, are

associated with oestrogen receptor-positive breast tumours and interact with metastasis gene C4.4a and dystroglycan. Br J Cancer 2003 Feb 24;88(4):579-85.

(173) Krig SR, Jin VX, Bieda MC, O'Geen H, Yaswen P, Green R, et al. Identification of genes directly regulated by the oncogene ZNF217 using chromatin immunoprecipitation (ChIP)-chip assays. J Biol Chem 2007 Mar 30;282(13):9703-12.

(174) Hengel SM, Murray E, Langdon S, Hayward L, O'Donoghue J, Panchaud A, et al. Data-independent proteomic screen identifies novel tamoxifen agonist that mediates drug resistance. J Proteome Res 2011 Oct 7;10(10):4567-78.

(175) Hrstka R, Nenutil R, Fourtouna A, Maslon MM, Naughton C, Langdon S, et al. The pro-metastatic protein anterior gradient-2 predicts poor prognosis in tamoxifen-treated breast cancers. Oncogene 2010 Aug 26;29(34):4838-47.

(176) Thompson DA, Weigel RJ. hAG-2, the human homologue of the Xenopus laevis cement gland gene XAG-2, is coexpressed with estrogen receptor in breast cancer cell lines. Biochem Biophys Res Commun 1998 Oct 9;251(1):111-6.

(177) Zhang JS, Gong A, Cheville JC, Smith DI, Young CY. AGR2, an androgen-inducible secretory protein overexpressed in prostate cancer. Genes Chromosomes Cancer 2005 Jul;43(3):249-59.

(178) Zweitzig DR, Smirnov DA, Connelly MC, Terstappen LW, O'Hara SM, Moran E. Physiological stress induces the metastasis marker AGR2 in breast cancer cells. Mol Cell Biochem 2007 Dec;306(1-2):255-60.

(179) Higa A, Mulot A, Delom F, Bouchecareilh M, Nguyen DT, Boismenu D, et al. Role of pro-oncogenic protein disulfide isomerase (PDI) family member anterior gradient 2 (AGR2) in the control of endoplasmic reticulum homeostasis. J Biol Chem 2011 Dec 30;286(52):44855-68.

(180) Norris AM, Gore A, Balboni A, Young A, Longnecker DS, Korc M. AGR2 is a SMAD4-suppressible gene that modulates MUC1 levels and promotes the initiation and progression of pancreatic intraepithelial neoplasia. Oncogene 2012 Sep 3.

(181) Myung JK, Frischer T, Afjehi-Sadat L, Pollak A, Lubec G. Mass spectrometrical analysis of the processed metastasis-inducing anterior gradient protein 2 homolog reveals 100% sequence coverage. Amino Acids 2008 Aug;35(2):485-94.

(182) Ward JJ, McGuffin LJ, Bryson K, Buxton BF, Jones DT. The DISOPRED server for the prediction of protein disorder. Bioinformatics 2004 Sep 1;20(13):2138-9.

(183) Iakoucheva LM, Brown CJ, Lawson JD, Obradovic Z, Dunker AK. Intrinsic disorder in cell-signaling and cancer-associated proteins. J Mol Biol 2002 Oct 25;323(3):573-84.

(184) Adam PJ, Boyd R, Tyson KL, Fletcher GC, Stamps A, Hudson L, et al. Comprehensive proteomic analysis of breast cancer cell membranes reveals unique proteins with potential roles in clinical cancer. J Biol Chem 2003 Feb 21;278(8):6482-9.

(185) Zhang Z, Henzel WJ. Signal peptide prediction based on analysis of experimentally verified cleavage sites. Protein Sci 2004 Oct;13(10):2819-24.

(186) Vitale A, Denecke J. The endoplasmic reticulum-gateway of the secretory pathway. Plant Cell 1999 Apr;11(4):615-28.

(187) Fourtouna A, Murray E, Nicholson J, Maslon MM, Pang L, Dryden DTF, et al. The anterior gradient-2 pathway as a model for developing peptide-aptamer anti-cancer drug leads that stimulate p53 function. Current Chemical Biology 2009 May 1;3:124-37.

(188) Gupta A, Dong A, Lowe AW. AGR2 gene function requires a unique endoplasmic reticulum localization motif. J Biol Chem 2012 Feb 10;287(7):4773-82.

(189) Munro S, Pelham HR. A C-terminal signal prevents secretion of luminal ER proteins. Cell 1987 Mar 13;48(5):899-907.

(190) Yamamoto K, Hamada H, Shinkai H, Kohno Y, Koseki H, Aoe T. The KDEL receptor modulates the endoplasmic reticulum stress response through mitogen-activated protein kinase signaling cascades. J Biol Chem 2003 Sep 5;278(36):34525-32.

(191) Gold LI, Eggleton P, Sweetwyne MT, Van Duyn LB, Greives MR, Naylor SM, et al. Calreticulin: non-endoplasmic reticulum functions in physiology and disease. FASEB J 2010 Mar;24(3):665-83.

(192) Raykhel I, Alanen H, Salo K, Jurvansuu J, Nguyen VD, Latva-Ranta M, et al. A molecular specificity code for the three mammalian KDEL receptors. J Cell Biol 2007 Dec 17;179(6):1193-204.

(193) Alanen HI, Raykhel IB, Luukas MJ, Salo KE, Ruddock LW. Beyond KDEL: the role of positions 5 and 6 in determining ER localization. J Mol Biol 2011 Jun 10;409(3):291-7.

(194) Dong A, Gupta A, Pai RK, Tun M, Lowe AW. The human adenocarcinoma-associated gene, AGR2, induces expression of amphiregulin through Hippo pathway co-activator YAP1 activation. J Biol Chem 2011 May 20;286(20):18301-10.

(195) Ferrari DM, Soling HD. The protein disulphide-isomerase family: unravelling a string of folds. Biochem J 1999 Apr 1;339 (Pt 1):1-10.

(196) Jessop CE, Watkins RH, Simmons JJ, Tasab M, Bulleid NJ. Protein disulphide isomerase family members show distinct substrate specificity: P5 is targeted to BiP client proteins. J Cell Sci 2009 Dec 1;122(Pt 23):4287-95.

(197) Anelli T, Alessio M, Mezghrani A, Simmen T, Talamo F, Bachi A, et al. ERp44, a novel endoplasmic reticulum folding assistant of the thioredoxin family. EMBO J 2002 Feb 15;21(4):835-44.

(198) Tachibana C, Stevens TH. The yeast EUG1 gene encodes an endoplasmic reticulum protein that is functionally related to protein disulfide isomerase. Mol Cell Biol 1992 Oct;12(10):4601-11.

(199) Maslon MM, Hrstka R, Vojtesek B, Hupp TR. A divergent substrate-binding loop within the pro-oncogenic protein anterior gradient-2 forms a docking site for Reptin. J Mol Biol 2010 Dec 3;404(3):418-38.

(200) Rowe ML, Ruddock LW, Kelly G, Schmidt JM, Williamson RA, Howard MJ. Solution structure and dynamics of ERp18, a small endoplasmic reticulum resident oxidoreductase. Biochemistry 2009 Jun 2;48(21):4596-606.

(201) Patel P, Clarke C, Barraclough DL, Jowitt TA, Rudland PS, Barraclough R, et al. Metastasis-promoting anterior gradient 2 protein has a dimeric thioredoxin fold structure and a role in cell adhesion. J Mol Biol 2013 Mar 11;425(5):929-43.

(202) Ryu J, Park SG, Lee PY, Cho S, Lee dH, Kim GH, et al. Dimerization of pro-oncogenic protein Anterior Gradient 2 is required for the interaction with BiP/GRP78. Biochem Biophys Res Commun 2013 Jan 11;430(2):610-5.

(203) Murray E. Identifying AGR2 interacting partners by covalent crosslinking coupled to mass spectrometry. Manuscript in preparation 2013.

(204) Dumartin L, Whiteman HJ, Weeks ME, Hariharan D, Dmitrovic B, Iacobuzio-Donahue CA, et al. AGR2 is a novel surface antigen that promotes the dissemination of pancreatic cancer cells through regulation of cathepsins B and D. Cancer Res 2011 Nov 15;71(22):7091-102.

(205) Brychtova V, Vojtesek B, Hrstka R. Anterior gradient 2: a novel player in tumor cell biology. Cancer Lett 2011 May 1;304(1):1-7.

(206) Vivekanandan P, Micchelli ST, Torbenson M. Anterior gradient-2 is overexpressed by fibrolamellar carcinomas. Hum Pathol 2009 Mar;40(3):293-9.

(207) Lepreux S, Bioulac-Sage P, Chevet E. Differential expression of the anterior gradient protein-2 is a conserved feature during morphogenesis and carcinogenesis of the biliary tree. Liver Int 2011 Mar;31(3):322-8.

(208) Yu H, Zhao J, Lin L, Zhang Y, Zhong F, Liu Y, et al. Proteomic study explores AGR2 as pro-metastatic protein in HCC. Mol Biosyst 2012 Jul 25.

(209) Persson S, Rosenquist M, Knoblach B, Khosravi-Far R, Sommarin M, Michalak M. Diversity of the protein disulfide isomerase family: identification of breast tumor induced Hag2 and Hag3 as novel members of the protein family. Mol Phylogenet Evol 2005 Sep;36(3):734-40.

(210) Chevet E, Cameron PH, Pelletier MF, Thomas DY, Bergeron JJ. The endoplasmic reticulum: integration of protein folding, quality control, signaling and degradation. Curr Opin Struct Biol 2001 Feb;11(1):120-4.

(211) Sevier CS, Kaiser CA. Formation and transfer of disulphide bonds in living cells. Nat Rev Mol Cell Biol 2002 Nov;3(11):836-47.

(212) Martin JL. Thioredoxin--a fold for all reasons. Structure 1995 Mar 15;3(3):245-50.

(213) Alanen HI, Williamson RA, Howard MJ, Lappi AK, Jantti HP, Rautio SM, et al. Functional characterization of ERp18, a new endoplasmic reticulum-located thioredoxin superfamily member. J Biol Chem 2003 Aug 1;278(31):28912-20.

(214) Cunnea PM, Miranda-Vizuete A, Bertoli G, Simmen T, Damdimopoulos AE, Hermann S, et al. ERdj5, an endoplasmic reticulum (ER)-resident protein containing DnaJ and thioredoxin domains, is expressed in secretory cells or following ER stress. J Biol Chem 2003 Jan 10;278(2):1059-66.

(215) Norgaard P, Winther JR. Mutation of yeast Eug1p CXXS active sites to CXXC results in a dramatic increase in protein disulphide isomerase activity. Biochem J 2001 Aug 15;358(Pt 1):269-74.

(216) Wunderlich M, Otto A, Maskos K, Mucke M, Seckler R, Glockshuber R. Efficient catalysis of disulfide formation during protein folding with a single active-site cysteine. J Mol Biol 1995 Mar 17;247(1):28-33.

(217) Li J, Lee B, Lee AS. Endoplasmic reticulum stress-induced apoptosis: multiple pathways and activation of p53-up-regulated modulator of apoptosis (PUMA) and NOXA by p53. J Biol Chem 2006 Mar 17;281(11):7260-70.

(218) Ramachandran V, Arumugam T, Wang H, Logsdon CD. Anterior gradient 2 is expressed and secreted during the development of pancreatic cancer and promotes cancer cell survival. Cancer Res 2008 Oct 1;68(19):7811-8.

(219) Liu D, Rudland PS, Sibson DR, Platt-Higgins A, Barraclough R. Human homologue of cement gland protein, a novel metastasis inducer associated with breast carcinomas. Cancer Res 2005 May 1;65(9):3796-805.

(220) Willmarth NE, Ethier SP. Autocrine and juxtacrine effects of amphiregulin on the proliferative, invasive, and migratory properties of normal and neoplastic human mammary epithelial cells. J Biol Chem 2006 Dec 8;281(49):37728-37.

(221) Ma L, Gauville C, Berthois Y, Millot G, Johnson GR, Calvo F. Antisense expression for amphiregulin suppresses tumorigenicity of a transformed human breast epithelial cell line. Oncogene 1999 Nov 11;18(47):6513-20.

(222) Shoyab M, Plowman GD, McDonald VL, Bradley JG, Todaro GJ. Structure and function of human amphiregulin: a member of the epidermal growth factor family. Science 1989 Feb 24;243(4894 Pt 1):1074-6.

(223) Herbst RS. Review of epidermal growth factor receptor biology. Int J Radiat Oncol Biol Phys 2004;59(2 Suppl):21-6.

(224) Zeng Q, Hong W. The emerging role of the hippo pathway in cell contact inhibition, organ size control, and cancer development in mammals. Cancer Cell 2008 Mar;13(3):188-92.

(225) Steinhardt AA, Gayyed MF, Klein AP, Dong J, Maitra A, Pan D, et al. Expression of Yes-associated protein in common solid tumors. Hum Pathol 2008 Nov;39(11):1582-9.

(226) Dong J, Feldmann G, Huang J, Wu S, Zhang N, Comerford SA, et al. Elucidation of a universal size-control mechanism in Drosophila and mammals. Cell 2007 Sep 21;130(6):1120-33.

(227) Kaiser J. Toxicology. Just how bad is dioxin? Science 2000 Jun 16;288(5473):1941-4.

(228) Paajarvi G, Jernstrom B, Seidel A, Stenius U. Anti-diol epoxide of benzo[a]pyrene induces transient Mdm2 and p53 Ser15 phosphorylation, while anti-diol epoxide of dibenzo[a,l]pyrene induces a nontransient p53 Ser15 phosphorylation. Mol Carcinog 2008 Apr;47(4):301-9.

(229) Paajarvi G, Viluksela M, Pohjanvirta R, Stenius U, Hogberg J. TCDD activates Mdm2 and attenuates the p53 response to DNA damaging agents. Carcinogenesis 2005 Jan;26(1):201-8.

(230) Ambolet-Camoit A, Bui LC, Pierre S, Chevallier A, Marchand A, Coumoul X, et al. 2,3,7,8-tetrachlorodibenzo-p-dioxin counteracts the p53 response to a genotoxicant by upregulating expression of the metastasis marker agr2 in the hepatocarcinoma cell line HepG2. Toxicol Sci 2010 Jun;115(2):501-12.

(231) Vanderlaag KE, Hudak S, Bald L, Fayadat-Dilman L, Sathe M, Grein J, et al. Anterior gradient-2 plays a critical role in breast cancer cell growth and survival by modulating cyclin D1, estrogen receptor-alpha and survivin. Breast Cancer Res 2010;12(3):R32.

(232) Di VE, Crahay C, Garbacki N, Hennuy B, Gueders M, Noel A, et al. New asthma biomarkers: lessons from murine models of acute and chronic asthma. Am J Physiol Lung Cell Mol Physiol 2009 Feb;296(2):L185-L197.

(233) Schroeder BW, Verhaeghe C, Park SW, Nguyenvu LT, Huang X, Zhen G, et al. AGR2 Is Induced in Asthma and Promotes Allergen-Induced Mucin Overproduction. Am J Respir Cell Mol Biol 2012 Aug;47(2):178-85.

(234) Kuyper LM, Pare PD, Hogg JC, Lambert RK, Ionescu D, Woods R, et al. Characterization of airway plugging in fatal asthma. Am J Med 2003 Jul;115(1):6-11.

(235) Ordonez CL, Khashayar R, Wong HH, Ferrando R, Wu R, Hyde DM, et al. Mild and moderate asthma is associated with airway goblet cell hyperplasia and abnormalities in mucin gene expression. Am J Respir Crit Care Med 2001 Feb;163(2):517-23.

(236) Yu H, Li Q, Kolosov VP, Perelman JM, Zhou X. Interleukin-13 induces mucin 5AC production involving STAT6/SPDEF in human airway epithelial cells. Cell Commun Adhes 2010 Aug;17(4-6):83-92.

(237) Zhou M, Chen HL, Cheng S, Mei L, Zhang HL, Xie M, et al. Effect of dexamethasone on expression of AGR2 protein in asthmatic mice. J Huazhong Univ Sci Technolog Med Sci 2013 Feb;33(1):33-6.

(238) Satsangi J, Parkes M, Louis E, Hashimoto L, Kato N, Welsh K, et al. Two stage genome-wide search in inflammatory bowel disease provides evidence for susceptibility loci on chromosomes 3, 7 and 12. Nat Genet 1996 Oct;14(2):199-202.

(239) Hampe J, Schreiber S, Shaw SH, Lau KF, Bridger S, MacPherson AJ, et al. A genomewide analysis provides evidence for novel linkages in inflammatory bowel disease in a large European cohort. Am J Hum Genet 1999 Mar;64(3):808-16.

(240) Bogaert S, De VM, Olievier K, Peeters H, Elewaut D, Lambrecht B, et al. Involvement of endoplasmic reticulum stress in inflammatory bowel disease: a different implication for colonic and ileal disease? PLoS One 2011;6(10):e25589.

(241) Barraclough DL, Platt-Higgins A, de Silva RS, Barraclough R, Winstanley J, West CR, et al. The metastasis-associated anterior gradient 2 protein is correlated with poor survival of breast cancer patients. Am J Pathol 2009 Nov;175(5):1848-57.

(242) Fritzsche FR, Dahl E, Pahl S, Burkhardt M, Luo J, Mayordomo E, et al. Prognostic relevance of AGR2 expression in breast cancer. Clin Cancer Res 2006 Mar 15;12(6):1728-34.

(243) Wu ZS, Wu Q, Ding XD, Wang HQ, Shen YX, Fang SY. [Expression of a novel metastasis-inducing protein human anterior gradient-2 (AGR2) in breast cancer and its clinical and prognostic significance]. Zhonghua Bing Li Xue Za Zhi 2008 Feb;37(2):109-13.

(244) Verma S, Salmans ML, Geyfman M, Wang H, Yu Z, Lu Z, et al. The estrogen-responsive Agr2 gene regulates mammary epithelial proliferation and facilitates lobuloalveolar development. Dev Biol 2012 Sep 15;369(2):249-60.

(245) Bu H, Bormann S, Schafer G, Horninger W, Massoner P, Neeb A, et al. The anterior gradient 2 (AGR2) gene is overexpressed in prostate cancer and may be useful as a urine sediment marker for prostate cancer detection. Prostate 2010 Oct 13.

(246) Zhang Y, Forootan SS, Liu D, Barraclough R, Foster CS, Rudland PS, et al. Increased expression of anterior gradient-2 is significantly associated with poor survival of prostate cancer patients. Prostate Cancer Prostatic Dis 2007;10(3):293-300.

(247) Edgell TA, Barraclough DL, Rajic A, Dhulia J, Lewis KJ, Armes JE, et al. Increased plasma concentrations of anterior gradient 2 protein are positively associated with ovarian cancer. Clin Sci (Lond) 2010 Jun;118(12):717-25.

(248) Gray TA, MacLaine NJ, Michie CO, Bouchalova P, Murray E, Howie J, et al. Anterior Gradient-3: a novel biomarker for ovarian cancer that mediates cisplatin resistance in xenograft models. J Immunol Methods 2012 Apr 30;378(1-2):20-32.

(249) Park K, Chung YJ, So H, Kim K, Park J, Oh M, et al. AGR2, a mucinous ovarian cancer marker, promotes cell proliferation and migration. Exp Mol Med 2011 Jan 4.

(250) Rice GE, Edgell TA, Autelitano DJ. Evaluation of midkine and anterior gradient 2 in a multimarker panel for the detection of ovarian cancer. J Exp Clin Cancer Res 2010;29:62.

(251) Yagui-Beltran A, Craig AL, Lawrie L, Thompson D, Pospisilova S, Johnston D, et al. The human oesophageal squamous epithelium exhibits a novel type of heat shock protein response. Eur J Biochem 2001 Oct;268(20):5343-55.

(252) Lee dH, Lee Y, Ryu J, Park SG, Cho S, Lee JJ, et al. Identification of proteins differentially expressed in gastric cancer cells with high metastatic potential for invasion to lymph nodes. Mol Cells 2011 Jun;31(6):563-71.

(253) Fritzsche FR, Dahl E, Dankof A, Burkhardt M, Pahl S, Petersen I, et al. Expression of AGR2 in non small cell lung cancer. Histol Histopathol 2007 Jul;22(7):703-8.

(254) Pizzi M, Fassan M, Realdon S, Balistreri M, Battaglia G, Giacometti C, et al. Anterior gradient 2 profiling in Barrett columnar epithelia and adenocarcinoma. Hum Pathol 2012 Nov;43(11):1839-44.

(255) Armes JE, Davies CM, Wallace S, Taheri T, Perrin LC, Autelitano DJ. AGR2 expression in ovarian tumours: a potential biomarker for endometrioid and mucinous differentiation. Pathology 2013 Jan;45(1):49-54.

(256) Darb-Esfahani S, Fritzsche F, Kristiansen G, Weichert W, Sehouli J, Braicu I, et al. Anterior gradient protein 2 (AGR2) is an independent prognostic factor in ovarian high-grade serous carcinoma. Virchows Arch 2012 Aug;461(2):109-16.

(257) Kristiansen G, Pilarsky C, Wissmann C, Kaiser S, Bruemmendorf T, Roepcke S, et al. Expression profiling of microdissected matched prostate cancer samples

reveals CD166/MEMD and CD24 as new prognostic markers for patient survival. J Pathol 2005 Feb;205(3):359-76.

(258) Kovalev LI, Shishkin SS, Khasigov PZ, Dzeranov NK, Kazachenko AV, Toropygin II, et al. [Identification of AGR2 protein, a novel potential cancer marker, using proteomics technologies]. Prikl Biokhim Mikrobiol 2006 Jul;42(4):480-4.

(259) Kani K, Malihi PD, Jiang Y, Wang H, Wang Y, Ruderman DL, et al. Anterior gradient 2 (AGR2): blood-based biomarker elevated in metastatic prostate cancer associated with the neuroendocrine phenotype. Prostate 2013 Feb 15;73(3):306-15.

(260) Zhang Y, Ali TZ, Zhou H, D'Souza DR, Lu Y, Jaffe J, et al. ErbB3 binding protein 1 represses metastasis-promoting gene anterior gradient protein 2 in prostate cancer. Cancer Res 2010 Jan 1;70(1):240-8.

(261) Zhang Y, Linn D, Liu Z, Melamed J, Tavora F, Young CY, et al. EBP1, an ErbB3-binding protein, is decreased in prostate cancer and implicated in hormone resistance. Mol Cancer Ther 2008 Oct;7(10):3176-86.

(262) Hu Z, Gu Y, Han B, Zhang J, Li Z, Tian K, et al. Knockdown of AGR2 induces cellular senescence in prostate cancer cells. Carcinogenesis 2012 Jun;33(6):1178-86.

(263) Chen R, Pan S, Duan X, Nelson BH, Sahota RA, de RS, et al. Elevated level of anterior gradient-2 in pancreatic juice from patients with pre-malignant pancreatic neoplasia. Mol Cancer 2010;9:149.

(264) Makawita S, Smith C, Batruch I, Zheng Y, Ruckert F, Grutzmann R, et al. Integrated proteomic profiling of cell line conditioned media and pancreatic juice for the identification of pancreatic cancer biomarkers. Mol Cell Proteomics 2011 Oct;10(10):M111.

(265) Valladares-Ayerbes M, Blanco-Calvo M, Reboredo M, Lorenzo-Patino MJ, Iglesias-Diaz P, Haz M, et al. Evaluation of the Adenocarcinoma-Associated Gene

AGR2 and the Intestinal Stem Cell Marker LGR5 as Biomarkers in Colorectal Cancer. Int J Mol Sci 2012;13(4):4367-87.

(266) Chung K, Nishiyama N, Yamano S, Komatsu H, Hanada S, Wei M, et al. Serum AGR2 as an early diagnostic and postoperative prognostic biomarker of human lung adenocarcinoma. Cancer Biomark 2011;10(2):101-7.

(267) Chung K, Nishiyama N, Wanibuchi H, Yamano S, Hanada S, Wei M, et al. AGR2 as a potential biomarker of human lung adenocarcinoma. Osaka City Med J 2012 Jun;58(1):13-24.

(268) Pizzi M, Fassan M, Balistreri M, Galligioni A, Rea F, Rugge M. Anterior gradient 2 overexpression in lung adenocarcinoma. Appl Immunohistochem Mol Morphol 2012 Jan;20(1):31-6.

(269) Missiaglia E, Blaveri E, Terris B, Wang YH, Costello E, Neoptolemos JP, et al. Analysis of gene expression in cancer cell lines identifies candidate markers for pancreatic tumorigenesis and metastasis. Int J Cancer 2004 Oct 20;112(1):100-12.

(270) Sweeny L, Liu Z, Bush BD, Hartman Y, Zhou T, Rosenthal EL. CD147 and AGR2 expression promote cellular proliferation and metastasis of head and neck squamous cell carcinoma. Exp Cell Res 2012 Aug 15;318(14):1788-98.

(271) Vendrell JA, Robertson KE, Ravel P, Bray SE, Bajard A, Purdie CA, et al. A candidate molecular signature associated with tamoxifen failure in primary breast cancer. Breast Cancer Res 2008;10(5):R88.

(272) Hrstka R, Murray E, Brychtova V, Fabian P, Hupp TR, Vojtesek B. Identification of an AKT-dependent signalling pathway that mediates tamoxifen-dependent induction of the pro-metastatic protein anterior gradient-2. Cancer Lett 2013 Jan 24.

(273) Huber M, Bahr I, Kratzschmar JR, Becker A, Muller EC, Donner P, et al. Comparison of proteomic and genomic analyses of the human breast cancer cell line

T47D and the antiestrogen-resistant derivative T47D-r. Mol Cell Proteomics 2004 Jan;3(1):43-55.

(274) Zhao L, Lee BY, Brown DA, Molloy MP, Marx GM, Pavlakis N, et al. Identification of candidate biomarkers of therapeutic response to docetaxel by proteomic profiling. Cancer Res 2009 Oct 1;69(19):7696-703.

(275) Murray E, McKenna EO, Burch LR, Dillon J, Langridge-Smith P, Kolch W, et al. Microarray-formatted clinical biomarker assay development using peptide aptamers to anterior gradient-2. Biochemistry 2007 Dec 4;46(48):13742-51.

(276) Bu H, Schweiger MR, Manke T, Wunderlich A, Timmermann B, Kerick M, et al. Anterior gradient 2 and 3--two prototype androgen-responsive genes transcriptionally upregulated by androgens and by oestrogens in prostate cancer cells. FEBS J 2013 Mar;280(5):1249-66.

(277) Schagger H, von JG. Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. Anal Biochem 1987 Nov 1;166(2):368-79.

(278) Laemmli UK. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 1970 Aug 15;227(5259):680-5.

(279) Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods 2001 Dec;25(4):402-8.

(280) Ihaka R, Gentleman R. R: a language for data analysis and graphics. Journal of Computation and Graphical Studies 1996;5:299-314.

(281) Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, et al. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol 2004;5(10):R80.

(282) Du P, Kibbe WA, Lin SM. lumi: a pipeline for processing Illumina microarray. Bioinformatics 2008 Jul 1;24(13):1547-8.

(283) Breitling R, Armengaud P, Amtmann A, Herzyk P. Rank products: a simple, yet powerful, new method to detect differentially regulated genes in replicated microarray experiments. FEBS Lett 2004 Aug 27;573(1-3):83-92.

(284) Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci U S A 1998 Dec 8;95(25):14863-8.

(285) Huang dW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 2009;4(1):44-57.

(286) Liu H, Naismith JH. A simple and efficient expression and purification system using two newly constructed vectors. Protein Expr Purif 2009 Feb;63(2):102-11.

(287) Ong SE, Mann M. A practical recipe for stable isotope labeling by amino acids in cell culture (SILAC). Nat Protoc 2006;1(6):2650-60.

(288) Shevchenko A, Wilm M, Vorm O, Mann M. Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. Anal Chem 1996 Mar 1;68(5):850-8.

(289) Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol 2008 Dec;26(12):1367-72.

(290) Finishing the euchromatic sequence of the human genome. Nature 2004 Oct 21;431(7011):931-45.

(291) Pruitt KD, Tatusova T, Maglott DR. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res 2007 Jan;35(Database issue):D61-D65.

(292) Cusick ME, Klitgord N, Vidal M, Hill DE. Interactome: gateway into systems biology. Hum Mol Genet 2005 Oct 15;14 Spec No. 2:R171-R181.

(293) Rual JF, Venkatesan K, Hao T, Hirozane-Kishikawa T, Dricot A, Li N, et al. Towards a proteome-scale map of the human protein-protein interaction network. Nature 2005 Oct 20;437(7062):1173-8.

(294) Jensen ON. Modification-specific proteomics: characterization of post-translational modifications by mass spectrometry. Curr Opin Chem Biol 2004 Feb;8(1):33-41.

(295) Ruland J, Sirard C, Elia A, MacPherson D, Wakeham A, Li L, et al. p53 accumulation, defective cell proliferation, and early embryonic lethality in mice lacking tsg101. Proc Natl Acad Sci U S A 2001 Feb 13;98(4):1859-64.

(296) Ito T, Chiba T, Ozawa R, Yoshida M, Hattori M, Sakaki Y. A comprehensive two-hybrid analysis to explore the yeast protein interactome. Proc Natl Acad Sci U S A 2001 Apr 10;98(8):4569-74.

(297) Blaydes JP, Craig AL, Wallace M, Ball HM, Traynor NJ, Gibbs NK, et al. Synergistic activation of p53-dependent transcription by two cooperating damage recognition pathways. Oncogene 2000 Aug 10;19(34):3829-39.

(298) Taura M, Eguma A, Suico MA, Shuto T, Koga T, Komatsu K, et al. p53 regulates Toll-like receptor 3 expression and function in human epithelial cell lines. Mol Cell Biol 2008 Nov;28(21):6557-67.

(299) Gronostajski RM, Sadowski PD. Determination of DNA sequences essential for FLP-mediated recombination by a novel method. J Biol Chem 1985 Oct 5;260(22):12320-7.

(300) Jayaram M. Two-micrometer circle site-specific recombination: the minimal substrate and the possible role of flanking sequences. Proc Natl Acad Sci U S A 1985 Sep;82(17):5875-9.

(301) Senecoff JF, Bruckner RC, Cox MM. The FLP recombinase of the yeast 2-micron plasmid: characterization of its recombination site. Proc Natl Acad Sci U S A 1985 Nov;82(21):7270-4.

(302) Fortier ME, Kent S, Ashdown H, Poole S, Boksa P, Luheshi GN. The viral mimic, polyinosinic:polycytidylic acid, induces fever in rats via an interleukin-1-dependent mechanism. Am J Physiol Regul Integr Comp Physiol 2004 Oct;287(4):R759-R766.

(303) Liang CC, Park AY, Guan JL. In vitro scratch assay: a convenient and inexpensive method for analysis of cell migration in vitro. Nat Protoc 2007;2(2):329-33.

(304) Geback T, Schulz MM, Koumoutsakos P, Detmar M. TScratch: a novel and simple software tool for automated analysis of monolayer wound healing assays. Biotechniques 2009 Apr;46(4):265-74.

(305) Ke N, Wang X, Xu X, Abassi YA. The xCELLigence system for real-time and label-free monitoring of cell viability. Methods Mol Biol 2011;740:33-43.

(306) Scanlan MJ, Simpson AJ, Old LJ. The cancer/testis genes: review, standardization, and commentary. Cancer Immun 2004 Jan 23;4:1.

(307) Murphy D. Gene expression studies using microarrays: principles, problems, and prospects. Adv Physiol Educ 2002 Dec;26(1-4):256-70.

(308) De BO, Arden KC, Boretti M, Vantomme V, De SC, Czekay S, et al. Characterization of the GAGE genes that are expressed in various human cancers and in normal testis. Cancer Res 1999 Jul 1;59(13):3157-65.

(309) Cilensek ZM, Yehiely F, Kular RK, Deiss LP. A member of the GAGE family of tumor antigens is an anti-apoptotic gene that confers resistance to Fas/CD95/APO-1, Interferon-gamma, taxol and gamma-irradiation. Cancer Biol Ther 2002 Jul;1(4):380-7.

(310) Simpson AJ, Caballero OL, Jungbluth A, Chen YT, Old LJ. Cancer/testis antigens, gametogenesis and cancer. Nat Rev Cancer 2005 Aug;5(8):615-25.

(311) Shapiro JA. Revisiting the central dogma in the 21st century. Ann N Y Acad Sci 2009 Oct;1178:6-28.

(312) Rogers S, Girolami M, Kolch W, Waters KM, Liu T, Thrall B, et al. Investigating the correspondence between transcriptomic and proteomic expression profiles using coupled cluster models. Bioinformatics 2008 Dec 15;24(24):2894-900.

(313) Dhingra V, Gupta M, Andacht T, Fu ZF. New frontiers in proteomics research: a perspective. Int J Pharm 2005 Aug 11;299(1-2):1-18.

(314) de Sousa AR, Penalva LO, Marcotte EM, Vogel C. Global signatures of protein and mRNA expression levels. Mol Biosyst 2009 Dec;5(12):1512-26.

(315) Vogel C, Marcotte EM. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. Nat Rev Genet 2012 Apr;13(4):227-32.

(316) Wilkins MR, Pasquali C, Appel RD, Ou K, Golaz O, Sanchez JC, et al. From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis. Biotechnology (N Y) 1996 Jan;14(1):61-5.

(317) Cohen AA, Geva-Zatorsky N, Eden E, Frenkel-Morgenstern M, Issaeva I, Sigal A, et al. Dynamic proteomics of individual cancer cells in response to a drug. Science 2008 Dec 5;322(5907):1511-6.

(318) Zolg JW, Langen H. How industry is approaching the search for new diagnostic markers and biomarkers. Mol Cell Proteomics 2004 Apr;3(4):345-54.

(319) Aebersold R, Goodlett DR. Mass spectrometry in proteomics. Chem Rev 2001 Feb;101(2):269-95.

(320) Fenn JB, Mann M, Meng CK, Wong SF, Whitehouse CM. Electrospray ionization for mass spectrometry of large biomolecules. Science 1989 Oct 6;246(4926):64-71.

(321) Hillenkamp F, Karas M, Beavis RC, Chait BT. Matrix-assisted laser desorption/ionization mass spectrometry of biopolymers. Anal Chem 1991 Dec 15;63(24):1193A-203A.

(322) Mann M, Wilm M. Error-tolerant identification of peptides in sequence databases by peptide sequence tags. Anal Chem 1994 Dec 15;66(24):4390-9.

(323) Johnson RS, Martin SA, Biemann K, Stults JT, Watson JT. Novel fragmentation process of peptides by collision-induced decomposition in a tandem mass spectrometer: differentiation of leucine and isoleucine. Anal Chem 1987 Nov 1;59(21):2621-5.

(324) Ong SE, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, et al. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. Mol Cell Proteomics 2002 May;1(5):376-86.

(325) Ong SE, Mann M. Stable isotope labeling by amino acids in cell culture for quantitative proteomics. Methods Mol Biol 2007;359:37-52.

(326) Ong SE, Foster LJ, Mann M. Mass spectrometric-based approaches in quantitative proteomics. Methods 2003 Feb;29(2):124-30.

(327) Burkhart JM, Schumbrutzki C, Wortelkamp S, Sickmann A, Zahedi RP. Systematic and quantitative comparison of digest efficiency and specificity reveals the impact of trypsin quality on MS-based proteomics. J Proteomics 2012 Feb 2;75(4):1454-62.

(328) Walsh GM, Rogalski JC, Klockenbusch C, Kast J. Mass spectrometry-based proteomics in biomedical research: emerging technologies and future strategies. Expert Rev Mol Med 2010;12:e30.

(329) Rajcevic U, Niclou SP, Jimenez CR. Proteomics strategies for target identification and biomarker discovery in cancer. Front Biosci 2009;14:3292-303.

(330) Bantscheff M, Kuster B. Quantitative mass spectrometry in proteomics. Anal Bioanal Chem 2012 Sep;404(4):937-8.

(331) Mueller LN, Brusniak MY, Mani DR, Aebersold R. An assessment of software solutions for the analysis of mass spectrometry based quantitative proteomics data. J Proteome Res 2008 Jan;7(1):51-61.

(332) Hart JR, Liao L, Ueno L, Yates JR, III, Vogt PK. Protein expression profiles of C3H 10T1/2 murine fibroblasts and of isogenic cells transformed by the H1047R mutant of phosphoinositide 3-kinase (PI3K). Cell Cycle 2011 Mar 15;10(6):971-6.

(333) Lund R, Leth-Larsen R, Jensen ON, Ditzel HJ. Efficient isolation and quantitative proteomic analysis of cancer cell plasma membrane proteins for identification of metastasis-associated cell surface markers. J Proteome Res 2009 Jun;8(6):3078-90.

(334) Bessarabova M, Ishkin A, JeBailey L, Nikolskaya T, Nikolsky Y. Knowledge-based analysis of proteomics data. BMC Bioinformatics 2012;13 Suppl 16:S13.

(335) Brennan PA, Jing J, Ethunandan M, Gorecki D. Dystroglycan complex in cancer. Eur J Surg Oncol 2004 Aug;30(6):589-92.

(336) Scholzen T, Gerdes J. The Ki-67 protein: from the known and the unknown. J Cell Physiol 2000 Mar;182(3):311-22.

(337) Endl E, Gerdes J. The Ki-67 protein: fascinating forms and an unknown function. Exp Cell Res 2000 Jun 15;257(2):231-7.

(338) Li L, Liao J, Ruland J, Mak TW, Cohen SN. A TSG101/MDM2 regulatory loop modulates MDM2 degradation and MDM2/p53 feedback control. Proc Natl Acad Sci U S A 2001 Feb 13;98(4):1619-24.

(339) Liu RT, Huang CC, You HL, Chou FF, Hu CC, Chao FP, et al. Overexpression of tumor susceptibility gene TSG101 in human papillary thyroid carcinomas. Oncogene 2002 Jul 18;21(31):4830-7.

(340) Jiang Y, Ou Y, Cheng X. Role of TSG101 in cancer. Front Biosci 2013;18:279-88.

(341) Zhu G, Gilchrist R, Borley N, Chng HW, Morgan M, Marshall JF, et al. Reduction of TSG101 protein has a negative impact on tumor cell growth. Int J Cancer 2004 Apr 20;109(4):541-7.

(342) Zhong Q, Chen Y, Jones D, Lee WH. Perturbation of TSG101 protein affects cell cycle progression. Cancer Res 1998 Jul 1;58(13):2699-702.

(343) Ong SE, Mann M. Mass spectrometry-based proteomics turns quantitative. Nat Chem Biol 2005 Oct;1(5):252-62.

(344) Andersen JS, Lam YW, Leung AK, Ong SE, Lyon CE, Lamond AI, et al. Nucleolar proteome dynamics. Nature 2005 Jan 6;433(7021):77-83.

(345) Blagoev B, Ong SE, Kratchmarova I, Mann M. Temporal analysis of phosphotyrosine-dependent signaling networks by quantitative proteomics. Nat Biotechnol 2004 Sep;22(9):1139-45.

(346) Cuomo A, Bonaldi T. Systems biology "on-the-fly": SILAC-based quantitative proteomics and RNAi approach in Drosophila melanogaster. Methods Mol Biol 2010;662:59-78.

(347) Wu CC, MacCoss MJ, Howell KE, Matthews DE, Yates JR, III. Metabolic labeling of mammalian organisms with stable isotopes for quantitative proteomic analysis. Anal Chem 2004 Sep 1;76(17):4951-9.

(348) Zanivan S, Krueger M, Mann M. In vivo quantitative proteomics: the SILAC mouse. Methods Mol Biol 2012;757:435-50.

(349) Larance M, Bailly AP, Pourkarimi E, Hay RT, Buchanan G, Coulthurst S, et al. Stable-isotope labeling with amino acids in nematodes. Nat Methods 2011;8(10):849-51.

(350) Yao X, Freas A, Ramirez J, Demirev PA, Fenselau C. Proteolytic 18O labeling for comparative proteomics: model studies with two serotypes of adenovirus. Anal Chem 2001 Jul 1;73(13):2836-42.

(351) Miyagi M, Rao KC. Proteolytic 18O-labeling strategies for quantitative proteomics. Mass Spectrom Rev 2007 Jan;26(1):121-36.

(352) Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nat Biotechnol 1999 Oct;17(10):994-9.

(353) Ross PL, Huang YN, Marchese JN, Williamson B, Parker K, Hattan S, et al. Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. Mol Cell Proteomics 2004 Dec;3(12):1154-69.

(354) Thompson A, Schafer J, Kuhn K, Kienle S, Schwarz J, Schmidt G, et al. Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. Anal Chem 2003 Apr 15;75(8):1895-904.

(355) Carvalho PC, Hewel J, Barbosa VC, Yates JR, III. Identifying differences in protein expression levels by spectral counting and feature selection. Genet Mol Res 2008;7(2):342-56.

(356) Old WM, Meyer-Arendt K, Aveline-Wolf L, Pierce KG, Mendoza A, Sevinsky JR, et al. Comparison of label-free methods for quantifying human proteins by shotgun proteomics. Mol Cell Proteomics 2005 Oct;4(10):1487-502.

(357) Kirkpatrick DS, Gerber SA, Gygi SP. The absolute quantification strategy: a general procedure for the quantification of proteins and post-translational modifications. Methods 2005 Mar;35(3):265-73.

(358) Pratt JM, Petty J, Riba-Garcia I, Robertson DH, Gaskell SJ, Oliver SG, et al. Dynamics of protein turnover, a missing dimension in proteomics. Mol Cell Proteomics 2002 Aug;1(8):579-91.

(359) Gu S, Pan S, Bradbury EM, Chen X. Precise peptide sequencing and protein quantification in the human proteome through in vivo lysine-specific mass tagging. J Am Soc Mass Spectrom 2003 Jan;14(1):1-7.

(360) Ong SE, Mittler G, Mann M. Identifying and quantifying in vivo methylation sites by heavy methyl SILAC. Nat Methods 2004 Nov;1(2):119-26.

(361) Scott L, Lamb J, Smith S, Wheatley DN. Single amino acid (arginine) deprivation: rapid and selective death of cultured transformed and malignant cells. Br J Cancer 2000 Sep;83(6):800-10.

(362) Wheatley DN, Scott L, Lamb J, Smith S. Single amino acid (arginine) restriction: growth and death of cultured HeLa and human diploid fibroblasts. Cell Physiol Biochem 2000;10(1-2):37-55.

(363) Ong SE, Kratchmarova I, Mann M. Properties of 13C-substituted arginine in stable isotope labeling by amino acids in cell culture (SILAC). J Proteome Res 2003 Mar;2(2):173-81.

(364) Gehrmann ML, Hathout Y, Fenselau C. Evaluation of metabolic labeling for comparative proteomics in breast cancer cells. J Proteome Res 2004 Sep;3(5):1063-8.

(365) Ginsberg D. E2F1 pathways to apoptosis. FEBS Lett 2002 Oct 2;529(1):122-5.

(366) Weber JD, Taylor LJ, Roussel MF, Sherr CJ, Bar-Sagi D. Nucleolar Arf sequesters Mdm2 and activates p53. Nat Cell Biol 1999 May;1(1):20-6.

(367) Ghazalpour A, Bennett B, Petyuk VA, Orozco L, Hagopian R, Mungrue IN, et al. Comparative analysis of proteome and transcriptome variation in mouse. PLoS Genet 2011 Jun;7(6):e1001393.

(368) Foss EJ, Radulovic D, Shaffer SA, Ruderfer DM, Bedalov A, Goodlett DR, et al. Genetic basis of proteome variation in yeast. Nat Genet 2007 Nov;39(11):1369-75.

(369) Babst M, Odorizzi G, Estepa EJ, Emr SD. Mammalian tumor susceptibility gene 101 (TSG101) and the yeast homologue, Vps23p, both function in late endosomal trafficking. Traffic 2000 Mar;1(3):248-58.

(370) Dupre S, Volland C, Haguenauer-Tsapis R. Membrane transport: ubiquitylation in endosomal sorting. Curr Biol 2001 Nov 13;11(22):R932-R934.

(371) Li L, Li X, Francke U, Cohen SN. The TSG101 tumor susceptibility gene is located in chromosome 11 band p15 and is mutated in human breast cancer. Cell 1997 Jan 10;88(1):143-54.

(372) Li L, Francke U, Cohen SN. Retraction. The TSG101 tumor susceptibility gene is located in chromosome 11 band p15 and is mutated in human breast cancer. Cell 1998 May 15;93(4):following.

(373) Feng GH, Lih CJ, Cohen SN. TSG101 protein steady-state level is regulated posttranslationally by an evolutionarily conserved COOH-terminal sequence. Cancer Res 2000 Mar 15;60(6):1736-41.

(374) Li L, Cohen SN. Tsg101: a novel tumor susceptibility gene isolated by controlled homozygous functional knockout of allelic loci in mammalian cells. Cell 1996 May 3;85(3):319-29.

(375) Branden C, Tooze J. Part I Basic Structural Principles. Introduction to Protein Structure. 2nd ed. Garland Science; 1998. p. 1-13.

(376) Morelli X, Hupp T. Searching for the Holy Grail; protein-protein interaction analysis and modulation. EMBO Rep 2012 Oct;13(10):877-9.

(377) Thangudu RR, Bryant SH, Panchenko AR, Madej T. Modulating protein-protein interactions with small molecules: the importance of binding hotspots. J Mol Biol 2012 Jan 13;415(2):443-53.

(378) Wright PE, Dyson HJ. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. J Mol Biol 1999 Oct 22;293(2):321-31.

(379) Tompa P, Fuxreiter M. Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions. Trends Biochem Sci 2008 Jan;33(1):2-8.

(380) Bonetta L. Protein-protein interactions: Interactome under construction. Nature 2010 Dec 9;468(7325):851-4.

(381) Ron D, Walter P. Signal integration in the endoplasmic reticulum unfolded protein response. Nat Rev Mol Cell Biol 2007 Jul;8(7):519-29.

(382) Rutkowski DT, Hegde RS. Regulation of basal cellular physiology by the homeostatic unfolded protein response. J Cell Biol 2010 May 31;189(5):783-94.

(383) Jeong W, Lee DY, Park S, Rhee SG. ERp16, an endoplasmic reticulum-resident thiol-disulfide oxidoreductase: biochemical properties and role in apoptosis induced by endoplasmic reticulum stress. J Biol Chem 2008 Sep 12;283(37):25557-66.

(384) Basse MJ, Betzi S, Bourgeas R, Bouzidi S, Chetrit B, Hamon V, et al. 2P2Idb: a structural database dedicated to orthosteric modulation of protein-protein interactions. Nucleic Acids Res 2013 Jan;41(Database issue):D824-D827.

(385) May LT, Leach K, Sexton PM, Christopoulos A. Allosteric modulation of G protein-coupled receptors. Annu Rev Pharmacol Toxicol 2007;47:1-51.

(386) Hayouka Z, Rosenbluh J, Levin A, Loya S, Lebendiker M, Veprintsev D, et al. Inhibiting HIV-1 integrase by shifting its oligomerization equilibrium. Proc Natl Acad Sci U S A 2007 May 15;104(20):8316-21.

(387) Marianayagam NJ, Sunde M, Matthews JM. The power of two: protein dimerization in biology. Trends Biochem Sci 2004 Nov;29(11):618-25.

(388) Blazer LL, Neubig RR. Small molecule protein-protein interaction inhibitors as CNS therapeutic agents: current progress and future hurdles. Neuropsychopharmacology 2009 Jan;34(1):126-41.

(389) Anastasiou D, Yu Y, Israelsen WJ, Jiang JK, Boxer MB, Hong BS, et al. Pyruvate kinase M2 activators promote tetramer formation and suppress tumorigenesis. Nat Chem Biol 2012 Oct;8(10):839-47.

(390) Morgan HP, O'Reilly FJ, Wear MA, O'Neill JR, Fothergill-Gilmore LA, Hupp T, et al. M2 pyruvate kinase provides a mechanism for nutrient sensing and regulation of cell proliferation. Proc Natl Acad Sci U S A 2013 Apr 9;110(15):5881-6.

(391) Neduva V, Russell RB. Peptides mediating interaction networks: new leads at last. Curr Opin Biotechnol 2006 Oct;17(5):465-71.

(392) Nussinov R, Tsai CJ. Allostery in disease and in drug discovery. Cell 2013 Apr 11;153(2):293-305.

(393) Jaffe EK. Impact of quaternary structure dynamics on allosteric drug discovery. Curr Top Med Chem 2013 Jan 1;13(1):55-63.

(394) Tjernberg A, Markova N, Griffiths WJ, Hallen D. DMSO-related effects in protein characterization. J Biomol Screen 2006 Mar;11(2):131-7.

(395) Orlando RA, Gonzales AM, Royer RE, Deck LM, Vander Jagt DL. A chemical analog of curcumin as an improved inhibitor of amyloid Abeta oligomerization. PLoS One 2012;7(3):e31869.

(396) Levine H, III. Alzheimer's beta-peptide oligomer formation at physiologic concentrations. Anal Biochem 2004 Dec 1;335(1):81-90.

(397) Robson AF, Hupp TR, Lickiss F, Ball KL, Faulds K, Graham D. Nanosensing protein allostery using a bivalent mouse double minute two (MDM2) assay. Proc Natl Acad Sci U S A 2012 May 22;109(21):8073-8.

(398) Bouchecareilh M, Higa A, Fribourg S, Moenner M, Chevet E. Peptides derived from the bifunctional kinase/RNase enzyme IRE1alpha modulate IRE1alpha activity and protect cells from endoplasmic reticulum stress. FASEB J 2011 Sep;25(9):3115-29.

(399) Hilser VJ, Thompson EB. Intrinsic disorder as a mechanism to optimize allosteric coupling in proteins. Proc Natl Acad Sci U S A 2007 May 15;104(20):8311-5.

(400) Fuxreiter M, Tompa P, Simon I. Local structural disorder imparts plasticity on linear motifs. Bioinformatics 2007 Apr 15;23(8):950-6.

(401) Hegyi H, Schad E, Tompa P. Structural disorder promotes assembly of protein complexes. BMC Struct Biol 2007;7:65.

(402) Tompa P. Intrinsically disordered proteins: a 10-year recap. Trends Biochem Sci 2012 Dec;37(12):509-16.

(403) Morrison JL, Breitling R, Higham DJ, Gilbert DR. A lock-and-key model for protein-protein interactions. Bioinformatics 2006 Aug 15;22(16):2012-9.

(404) Trellet M, Melquiond AS, Bonvin AM. A unified conformational selection and induced fit approach to protein-peptide docking. PLoS One 2013;8(3):e58769.

(405) Wells JA, McClendon CL. Reaching for high-hanging fruit in drug discovery at protein-protein interfaces. Nature 2007 Dec 13;450(7172):1001-9.

(406) Liu J, Perumal NB, Oldfield CJ, Su EW, Uversky VN, Dunker AK. Intrinsic disorder in transcription factors. Biochemistry 2006 Jun 6;45(22):6873-88.

(407) Harvey AL. Natural products in drug discovery. Drug Discov Today 2008 Oct;13(19-20):894-901.

(408) Mondal S, Bandyopadhyay S, Ghosh MK, Mukhopadhyay S, Roy S, Mandal C. Natural products: promising resources for cancer drug discovery. Anticancer Agents Med Chem 2012 Jan;12(1):49-75.

(409) Ganesan A. The impact of natural products upon modern drug discovery. Curr Opin Chem Biol 2008 Jun;12(3):306-17.

(410) Harvey AL. Natural products as a screening resource. Curr Opin Chem Biol 2007 Oct;11(5):480-4.

(411) Acharya KR, Lloyd MD. The advantages and limitations of protein crystal structures. Trends Pharmacol Sci 2005 Jan;26(1):10-4.

(412) Heras B, Martin JL. Post-crystallization treatments for improving diffraction quality of protein crystals. Acta Crystallogr D Biol Crystallogr 2005 Sep;61(Pt 9):1173-80.

(413) Kleywegt GJ. Validation of protein crystal structures. Acta Crystallogr D Biol Crystallogr 2000 Mar;56(Pt 3):249-65.

(414) Chayen NE, Saridakis E. Protein crystallization: from purified protein to diffraction-quality crystal. Nat Methods 2008 Feb;5(2):147-53.

(415) Chayen NE. Turning protein crystallisation from an art into a science. Curr Opin Struct Biol 2004 Oct;14(5):577-83.

(416) Gorrec F. The MORPHEUS protein crystallization screen. J Appl Crystallogr 2009 Dec 1;42(Pt 6):1035-42.

(417) Nguyen VD, Ruddock LW, Salin M, Wierenga RK. Crystal struture of the human anterior gradient protein 3. 2013. Ref Type: Unpublished Work

(418) Ghosh S, Nie A, An J, Huang Z. Structure-based virtual screening of chemical libraries for drug discovery. Curr Opin Chem Biol 2006 Jun;10(3):194-202.

(419) Percy AJ, Rey M, Burns KM, Schriemer DC. Probing protein interactions with hydrogen/deuterium exchange and mass spectrometry-a review. Anal Chim Acta 2012 Apr 6;721:7-21.

(420) Innes HE, Liu D, Barraclough R, Davies MP, O'Neill PA, Platt-Higgins A, et al. Significance of the metastasis-inducing protein AGR2 for outcome in hormonally treated breast cancer patients. Br J Cancer 2006 Apr 10;94(7):1057-65.

(421) Meissner F, Scheltema RA, Mollenkopf HJ, Mann M. Direct proteomic quantification of the secretome of activated immune cells. Science 2013 Apr 26;340(6131):475-8.

(422) Barderas R, Mendes M, Torres S, Bartolome RA, Lopez-Lucendo M, Villar-Vazquez R, et al. In-depth characterization of the secretome of colorectal cancer metastatic cells identifies key proteins in cell adhesion, migration and invasion. Mol Cell Proteomics 2013 Feb 26.